Syst. Biol. 66(4):531–550, 2017
© The Author(s) 2016. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/syw069
Advance Access publication August 18, 2016

Disentangling Incomplete Lineage Sorting and Introgression to Refine Species-Tree Estimates for Lake Tanganyika Cichlid Fishes

Britta S. Meyer[1,2], Michael Matschiner[1,3], and Walter Salzburger[1,3,*]

[1]*Zoological Institute, University of Basel, Basel, Switzerland;* [2]*Evolutionary Ecology of Marine Fishes, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany;* [3]*Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway;*
**Correspondence to be sent to: Prof. Dr. Walter Salzburger, Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland E-mail: walter.salzburger@unibas.ch.*

Britta S. Meyer and Michael Matschiner contributed equally to this work.

Received 10 February 2016; reviews returned 16 June 2016; accepted 27 June 2016
Associate Editor: Laura Kubatko

Abstract.—Adaptive radiation is thought to be responsible for the evolution of a great portion of the past and present diversity of life. Instances of adaptive radiation, characterized by the rapid emergence of an array of species as a consequence to their adaptation to distinct ecological niches, are important study systems in evolutionary biology. However, because of the rapid lineage formation in these groups, and occasional gene flow between the participating species, it is often difficult to reconstruct the phylogenetic history of species that underwent an adaptive radiation. In this study, we present a novel approach for species-tree estimation in rapidly diversifying lineages, where introgression is known to occur, and apply it to a multimarker data set containing up to 16 specimens per species for a set of 45 species of East African cichlid fishes (522 individuals in total), with a main focus on the cichlid species flock of Lake Tanganyika. We first identified, using age distributions of most recent common ancestors in individual gene trees, those lineages in our data set that show strong signatures of past introgression. This led us to formulate three hypotheses of introgression between different lineages of Tanganyika cichlids: the ancestor of Boulengerochromini (or of Boulengerochromini and Bathybatini) received genomic material from the derived H-lineage; the common ancestor of Cyprichromini and Perissodini experienced, in turn, introgression from Boulengerochromini and/or Bathybatini; and the Lake Tanganyika Haplochromini and closely related riverine lineages received genetic material from Cyphotilapiini. We then applied the multispecies coalescent model to estimate the species tree of Lake Tanganyika cichlids, but excluded the lineages involved in these introgression events, as the multispecies coalescent model does not incorporate introgression. This resulted in a robust species tree, in which the Lamprologini were placed as sister lineage to the H-lineage (including the Eretmodini), and we identify a series of rapid splitting events at the base of the H-lineage. Divergence ages estimated with the multispecies coalescent model were substantially younger than age estimates based on concatenation, and agree with the geological history of the Great Lakes of East Africa. Finally, we formally tested the three hypotheses of introgression using a likelihood framework, and find strong support for introgression between some of the cichlid tribes of Lake Tanganyika. [Adaptive radiation; Cichlidae; introgression; Lake Tanganyika; species network.]

Adaptive radiation, that is, the rapid emergence of new life-forms through the extensive diversification of an organismal lineage into new or available ecological niches, is thought to be responsible for a great deal of the extant and extinct organismal diversity on our planet (Simpson 1953; Schluter 2000; Berner and Salzburger 2015). Rapid evolutionary radiations have long fascinated biologists and serve as model systems to explore mechanisms of lineage diversification (see, e.g., Schluter 2000; Mayr 2001; Coyne and Orr 2004). At the same time—mainly because of the rapidity of lineage formation facilitating incomplete lineage sorting, gene flow between the emerging species, and the frequent occurrence of phenotypic convergence—it has proven notoriously difficult to reconstruct the progression of adaptive radiations by means of morphological analyses (Fryer and Iles 1972) and, more recently, phylogenetic reconstructions based on molecular data (e.g., Rokas and Carroll 2006; Glor 2010; Meyer et al. 2015).

The retention of ancestral polymorphisms between species (i.e., incomplete lineage sorting) and gene exchange between taxa (i.e., introgressive hybridization) are commonly observed phenomena in adaptive radiations (see, e.g., The Heliconius Genome Consortium 2012; Brawand et al. 2014; Lamichhaney et al. 2015; reviewed in Berner and Salzburger 2015), and it has been suggested that extensive hybridization between founding lineages producing a so-called hybrid swarm may trigger subsequent adaptive radiation in the first place (Seehausen 2004; Nolte and Tautz 2010; Abbott et al. 2013). There are two major technical challenges for phylogeny reconstruction associated with the occurrence of incomplete lineage sorting and/or introgressive hybridization: First, both processes can generate phylogenetic histories that are discordant with the species tree, and can therefore cause misleading species-tree estimates unless properly accounted for (see, e.g., Maddison and Knowles 2006; Kubatko and Degnan 2007; Yu et al. 2011; Roch and Steel 2014; Edwards et al. 2016). Second, in order to account for incomplete lineage sorting and introgressive hybridization in phylogenetic analyses, it is important to differentiate between these two processes, since models used for phylogenetic inference often incorporate only one of the two processes (e.g., Linz et al. 2007; Heled and Drummond 2010; Mirarab et al. 2014). However, due to the various possible outcomes of both processes, it is often difficult to distinguish whether an observed incongruence between phylogenies is due to incomplete lineage sorting or past hybridization (Holder et al. 2001; Holland et al. 2008; Joly et al. 2009). As these issues are particularly pronounced when rapidly radiating

clades are involved, a thorough understanding of the evolutionary history of adaptive radiations is often hampered by the difficulties involved in phylogenetic analyses of these groups (Jarvis et al. 2014; Lamichhaney et al. 2015).

The most well-known examples of adaptive radiation include Darwin's finches on the Galapagos Islands (Grant and Grant 2008), threespine stickleback fish in the Northern Hemisphere (Bell and Foster 1994), anole lizards on the islands of the Caribbean (Losos 2009), and cichlid fishes in the East African Great Lakes (Fryer and Iles 1972). The exceptionally diverse cichlid assemblages in Lakes Victoria, Malawi, and Tanganyika represent the most species-rich extant adaptive radiations (Salzburger et al. 2014). Hundreds of closely related cichlid species have emerged in each of these lakes in the last few millions to several thousands of years (Kocher 2004; Seehausen 2006; Salzburger 2009; Salzburger et al. 2014), rendering the formation of new cichlid species in these lakes unusually rapid (McCune and Lovejoy 1998; Coyne and Orr 2004).

With an age of 9–12 myr, Lake Tanganyika is the oldest and, because of its great depth, also the most stable lake in Africa (Cohen et al. 1993; Salzburger et al. 2014). This has strong implications on its cichlid fauna, which is genetically, morphologically, ecologically, and behaviorally the most diverse of all extant cichlid species flocks (Salzburger et al. 2014). Lake Tanganyika is also a reservoir of more ancient lineages and the likely cradle of more modern cichlid groups (Nishida 1991; Salzburger et al. 2002a, 2005), or simply *the* "melting pot" of East African cichlid diversity (Weiss et al. 2015). Knowledge about the evolution of the Lake Tanganyika cichlid assemblage is thus key to understanding cichlid evolution in the whole of East Africa.

The cichlid fauna of Lake Tanganyika, which comprises about 200 species, has been divided into 14 taxonomic subgroups, so-called 'tribes' (i.e., a taxonomic rank between subfamily and genus) (Poll 1986; Dunz and Schliewen 2013; Takahashi and Sota 2016): Bathybatini, Boulengerochromini, Cyphotilapiini, Cyprichromini, Ectodini, Eretmodini, Haplochromini, Lamprologini, Limnochromini, Oreochromini, Perissodini, Tilapiini, Trematocarini, and Tylochromini. Over the past quarter of a century, a number of attempts have been undertaken to resolve the phylogenetic relationships between the cichlid tribes in Lake Tanganyika (e.g., Nishida 1991; Kocher et al. 1995; Salzburger et al. 2002a; Clabaut et al. 2005; Day et al. 2008; Muschick et al. 2012; Meyer et al. 2015; Weiss et al. 2015; McGee et al. 2016; Takahashi and Sota 2016), as well as between genera and species within tribes (see, e.g., Koblmüller et al. 2004, 2010; Sturmbauer et al. 2010). Especially with respect to the placement of tribes relative to each other, there is little consensus between the different studies (see Supplementary Text S1 and Supplementary Fig. S1; available on Dryad at http://dx.doi.org/10.5061/dryad.62qn5), which can in part be explained by the different marker sets (and phylogenetic methods) that were used at a given time. In addition, there is evidence for introgressive hybridization and incomplete lineage sorting in the course of the cichlid adaptive radiation in Lake Tanganyika (see, e.g., Rüber et al. 2001; Salzburger et al. 2002b; Koblmüller et al. 2010; Meyer et al. 2015; Weiss et al. 2015), and both of these processes may have contributed to the conflicting results of studies conducted with different markers. Overall, while the general phylogenetic structure of the cichlid assemblage of Lake Tanganyika is well established—with secondary colonizers of Tylochromini and Oreochromini as the earliest diverging lineages, and Trematocarini, Bathybatini, and Boulengerochromini as ancestral tribes that are sister to a clade formed by the Lamprologini and the H-lineage (Meyer et al. 1990; Nishida 1991; Day et al. 2008; Muschick et al. 2012; Weiss et al. 2015; Takahashi and Sota 2016)—several areas of uncertainty remain (see Meyer et al. 2015; Weiss et al. 2015; McGee et al. 2016; Takahashi and Sota 2016). For example, the relative position of Trematocarini, Bathybatini, and Boulengerochromini to one another is still unclear, as is the relative position of the H-lineage taxa to one another (in particular, the placement of Eretmodini).

In this study, we take a novel approach to address the phylogenetic relationships between the cichlid tribes of East African Lake Tanganyika. We use 40 nuclear markers established in Meyer and Salzburger (2012) and Meyer et al. (2015). Yet, instead of using a single representative per species as in Meyer et al. (2015), we use up to 16 specimens per species, resulting in nuclear DNA sequences from 522 specimens. This allows us to first disentangle the signals of introgression and incomplete lineage sorting among cichlid taxa, using new methodology to detect gene flow from population-level and species-level data sets. We then apply the multispecies coalescent model to identify the species tree of Lake Tanganyika cichlid lineages after filtering out signals of introgression that would represent a violation of this model. Finally, by comparing age estimates recovered with the multispecies coalescent model and with concatenation, we reconcile phylogenetic divergence-date estimates of cichlids with the geological history of their environment, to build a coherent timeline of cichlid diversification in the Great Lakes of East Africa, with a particular focus on Lake Tanganyika.

Materials and Methods

Data Collection and Sequence Alignment

All specimens of cichlid fishes from Lake Tanganyika used in this study were collected in accordance with the national legislation of the Republic of Zambia and under the memorandum of understanding (MOU) between the involved institutions (University of Basel, Switzerland, University of Zambia, Lusaka, Zambia, and Department of Fisheries, Lake Tanganyika branch, Mpulungu, Zambia). Fish collections were conducted

as described in Muschick et al. (2012). We used up to 16 individuals from 45 different cichlid species (a total of 522 individuals). Our sampling design comprised representatives of all major cichlid lineages of the Lake Tanganyika radiation (33 species) plus four representatives each of Lake Victoria and Lake Malawi, and four riverine taxa. The gross of samples originated from field trips to Zambia, that is, to Lake Tanganyika and the Kafue River in the years 2007, 2008, and 2011. Further samples were aquaria-bred at the Swiss Federal Institute of Aquatic Science and Technology (EAWAG; these were kindly provided by Ole Seehausen) and at the University of Basel. The number of sampled individuals per species and their place of origin are listed in Supplementary Tables S1 and S2 available on Dryad. Fin clips were stored at −20 °C in 100% ethanol until further processing.

Extraction of DNA from ethanol-preserved fin clips was performed with a Qiagen Biosprint 96 robot following the manufacturer's protocol (QIAGEN, Hombrechtikon, Switzerland). For multiplexed PCR amplifications, we used a primer set consisting of 40 markers (encompassing exons, introns, and untranslated regions of nuclear coding genes) as described in Meyer and Salzburger (2012) and Meyer et al. (2015) (see Table 1 for further details). PCR amplifications and library preparations (including several cleaning and pooling steps) were carried out as described in Meyer et al. (2015); pyrosequencing was performed on the GS FLX system (454 Sequencing, Roche). The sequencing design in the current study differs from that of Meyer et al. (2015), who included only a single specimen for each species (instead of up to 16 as in the current study) and used a consensus sequence (instead of the two alleles as in the current study).

The generated sequence reads were demultiplexed and assembled using Roche's sffinfo tool, PRINSEQ (Schmieder and Edwards 2011), BWA (Li and Durbin 2010), and Geneious (Biomatters Ltd, Auckland, New Zealand; available at www.geneious.com) with the same versions and settings as in Meyer et al. (2015). Due to the long read length provided by the 454 technology, phased alleles could be determined in all individuals. Thus, we obtained between 376 and 1036 (mean = 918.9) sequences per marker (Table 1). Alignments for each marker containing all alleles from all individuals were obtained with the software MAFFT v.7.017 (Katoh and Standley 2013), using the "–auto" option to automatically detect the appropriate alignment strategy. The alignments were inspected visually and locally corrected where necessary (Supplementary File S1 available on Dryad). To test whether different codon positions and noncoding fractions of the alignments are best characterized by different models of sequence evolution, we analyzed each alignment separately with the software PartitionFinder v.1.1.1 (Lanfear et al. 2012). However, as the program identified only one partition for each alignment, a single substitution model per marker was applied in all phylogenetic analyses.

Table 1. Description of the phylogenetic markers used in this study

Name (synonym)	Ensembl link	No. of individuals	No. of sites	Variable sites (%)	π
rag1	rag1	404	417	16.07	0.0114
b2m[a]	b2ml	448	479	22.13	0.0233
gapdhs	gapdhs	492	461	20.82	0.0086
ptchd4	PTR	483	393	11.70	0.0087
enc1	ENC1	516	372	8.60	0.0083
phpt1	novel gene	366	471	25.48	0.0141
rps7	rps7	455	390	32.31	0.0158
tbr1	TBR1	488	468	3.85	0.0039
aqp1a.1[a]	aqp1a.1	509	443	20.09	0.0128
hprt1[a]	hprt1	389	399	18.30	0.0117
anxa4	anxa4	491	732	13.11	0.0080
pgk1	pgk1	496	375	27.47	0.0109
bmp4	bmp4	514	465	17.63	0.0091
bmp2	bmp2b	479	372	13.98	0.0055
TMO-4C4	TTN (2 of 2)	512	428	17.52	0.0174
fgf6b	fgf6b (1 of 2)	384	470	8.72	0.0062
runx2	runx2b	499	362	6.91	0.0063
furina	furina	474	309	16.18	0.0104
wnt7b[b]	wnt7bb	516	387	6.46	0.0032
pax9	pax9	188	393	7.89	0.0047
sox10b	sox10 (1 of 2)	506	387	14.99	0.0098
otx2	OTX2 (2 of 2)	510	411	7.30	0.0033
otx1	otx1b	512	354	9.32	0.0033
dlx[a]	dlx2a	495	479	26.51	0.0136
dlx4b	dlx4b	480	355	14.93	0.0064
barx1	barx1	435	206	23.79	0.0119
ednrb1a	ednrba	504	443	22.35	0.0144
mc1r	mc1r	454	405	12.10	0.0074
skia	skia	475	450	10.22	0.0047
kita	kita	492	422	18.01	0.0103
mitfa	mitfa	383	435	21.15	0.0124
tyr	tyr	475	491	26.48	0.0204
hagoromo (fbxw4)	HAGOROMO	466	495	36.77	0.0414
slc45a2 (aim)	slc45a2	516	285	21.05	0.0137
rh1	RH1	518	402	13.18	0.0149
opn1mw (lws)	opn1lw2	469	411	21.41	0.0163
opn1sw (sws)	opn1sw1	282	430	23.95	0.0183
rpl13a[a]	rpl13a	464	368	17.39	0.0066
edar	edar	415	372	16.13	0.0107
csf1ra	csf1ra	489	368	21.20	0.0138

Notes: The marker name, the link to the corresponding Ensembl entry for Tilapia, the number of sequenced individuals, the alignment length of each marker, the proportion of variable sites, and the genetic diversity (π) of each marker are given. [a]Marker excluded from all gene-tree analyses and MRCA age comparisons due to missing data in *Tylochromis polylepis*. [b]Unconstrained gene-tree analyses failed for this marker, and it was thus excluded from MRCA age comparisons.

Gene-Tree Inference

As our phylogenetic data set included both intra- and interspecific sequence variation, we performed phylogenetic inference using the multispecies coalescent model of *BEAST (Heled and Drummond 2010), implemented in the Bayesian phylogenetic inference software BEAST v.2.1.3 (Bouckaert et al. 2014). However, we expected that gene tree topologies could be discordant not only due to incomplete lineage sorting, but also due to introgression. As the *BEAST model only accounts for incomplete lineage sorting but excludes introgression, we conducted separate *BEAST

analyses for each marker, assuming that recombination only takes place between markers, but not within. Phylogenetic inference was performed using an HKY model (Hasegawa et al. 1985) of sequence evolution and the Yule process (Yule 1925) as the species-tree prior. For computational reasons, we assumed no among-site rate heterogeneity and used empirical base frequencies. We also assumed a strict molecular clock model, to maximize comparability between divergence times in individual gene trees and to reduce the run time required for convergence. The molecular clock was calibrated with two fossil-based node-age constraints. Based on the time-calibrated phylogeny of cichlid fishes by McMahan et al. (2013), we constrained the divergence of Tylochromini and Austrotilapiini between 54.8 and 33.0 million years ago (Ma) (95% confidence interval; mean = 42.1 Ma), and that of Oreochromini and Austrotilapiini between 31.6 and 17.3 Ma (mean = 23.3 Ma). In our data set, Tylochromini were represented by *Tylochromis polylepis*, Oreochromini were represented by *Oreochromis tanganicae*, and the 43 remaining species were members of Austrotilapiini, which were assumed to be monophyletic. Both time constraints were implemented as lognormal prior distributions, with offsets 18.2 and 8.2 Ma, means (in real space) 23.894 and 15.067 Ma, and standard deviations 0.231 and 0.240 Ma, respectively. For consistency between ages estimated with different markers, we excluded five out of 40 markers due to missing sequence data for *T. polylepis* (see Table 1). We conducted five replicate analyses for each of the 35 markers, with 4 billion Markov chain Monte Carlo (MCMC) generations of which the first 10% were discarded as burn-in. These analyses were performed on the Maia cluster of the University of Basel and required about 40 CPU days per replicate. Run convergence was assessed by comparison of traces within and between run replicates and by effective sample sizes (ESS) greater than 200 for each model parameter. In only few exceptions, ESS values of single parameters remained below 200, but were still above 100. For each marker, 100 time-calibrated gene trees were sampled from the joint posterior distribution of the five replicate analyses. Each of these gene trees was further pruned to include only a single randomly chosen individual per species. This was repeated 10 times per tree to produce posterior sets of 1000 species-level gene trees for each marker. A flow chart illustrating our phylogenetic analysis pipeline is provided in Supplementary Fig. S2 available on Dryad.

Detection of Putatively Introgressed Lineages

Comparison of mean MRCA age estimates.—Posterior samples of species-level gene tree distributions were used to calculate mean age estimates of most recent common ancestors (MRCA) in all pairwise species comparisons. As gene-tree inference for one out of 35 markers (wnt7b) failed to converge, we used gene tree distributions of 34 markers for these calculations. Following Marcussen et al. (2014), we expected that introgression could be detectable through MRCA age estimates in three-taxon comparisons (see Fig. 1 for illustration): for a given set of taxa A, B, C, two of the three mean MRCA age estimates should be similar if introgression is absent. For example, if A and B are sister lineages and C is the outgroup then the MRCA for A and C and the MRCA for B and C should have a similar age, even in the presence of incomplete lineage sorting (assuming comparable population sizes and mutation rates for A and B). However, if part of the genome of A is affected by introgression coming from C, then the mean MRCA age for A and C should be younger than the MRCA for B and C. Thus, to identify the species with the strongest signals of introgression, we compared mean MRCA age estimates in each possible three-taxon combination, testing each of the positions A, B, and C for each of the 45 species in our data set (a total of 45^3 comparisons). We recorded the difference between the age estimate of the MRCA of B and C and the age estimate of the MRCA of A and C if: (i) the MRCA of A and C appeared younger than that of B and C, and if (ii) the MRCA of A and B was the youngest of the three, indicating that A and B are sister groups in this three-taxon comparison. If these conditions were not met, the age difference was recorded as zero. The results of these three-taxon comparisons were summarized by calculating, for each pair of the 45 species, the maximal age difference found with these two species in positions A and C, and with any other species in position B. For each of the 45 species, we also calculated the maximal age difference when this species was in position A, and any other two species were in positions B and C (thus the maximal age difference in 45^2 comparisons).

Based on the maximal age differences per species pair and per individual species, we developed three hypotheses of introgression between Lake Tanganyika cichlid lineages (Fig. 1; Supplementary Table S3 available on Dryad) that were to be tested in subsequent analyses: H1; an early representative of Boulengerochromini, or a common ancestor of Boulengerochromini and Bathybatini, received, through introgression, genomic material from members of the H-lineage, most likely from the Perissodini or Limnochromini; H2; an early member or common ancestor of the Perissodini and Cyprichromini experienced introgression from the Boulengerochromini or Bathybatini; H3; a common ancestor of the Lake Tanganyika Haplochromini (i.e., the "Tropheini" *sensu* Takahashi 2003) plus the riverine lineages *Pseudocrenilabrus* and *Serranochromis* experienced introgression from an ancestor or early member of Cyphotilapiini. A fourth potential introgression event was indicated by high maximal age differences in comparisons involving Haplochromini of lakes Malawi and Victoria (or one of *Astatotilapia burtoni* and *Haplochromis* sp. 'Chipwa') as taxon C and Perissodini as taxon A (see Fig. 1). However, as these signals for introgression between the Haplochromini and Perissodini were not as strong as those for introgression between the Boulengerochromini or Bathybatini and Perissodini (H2), we did not further

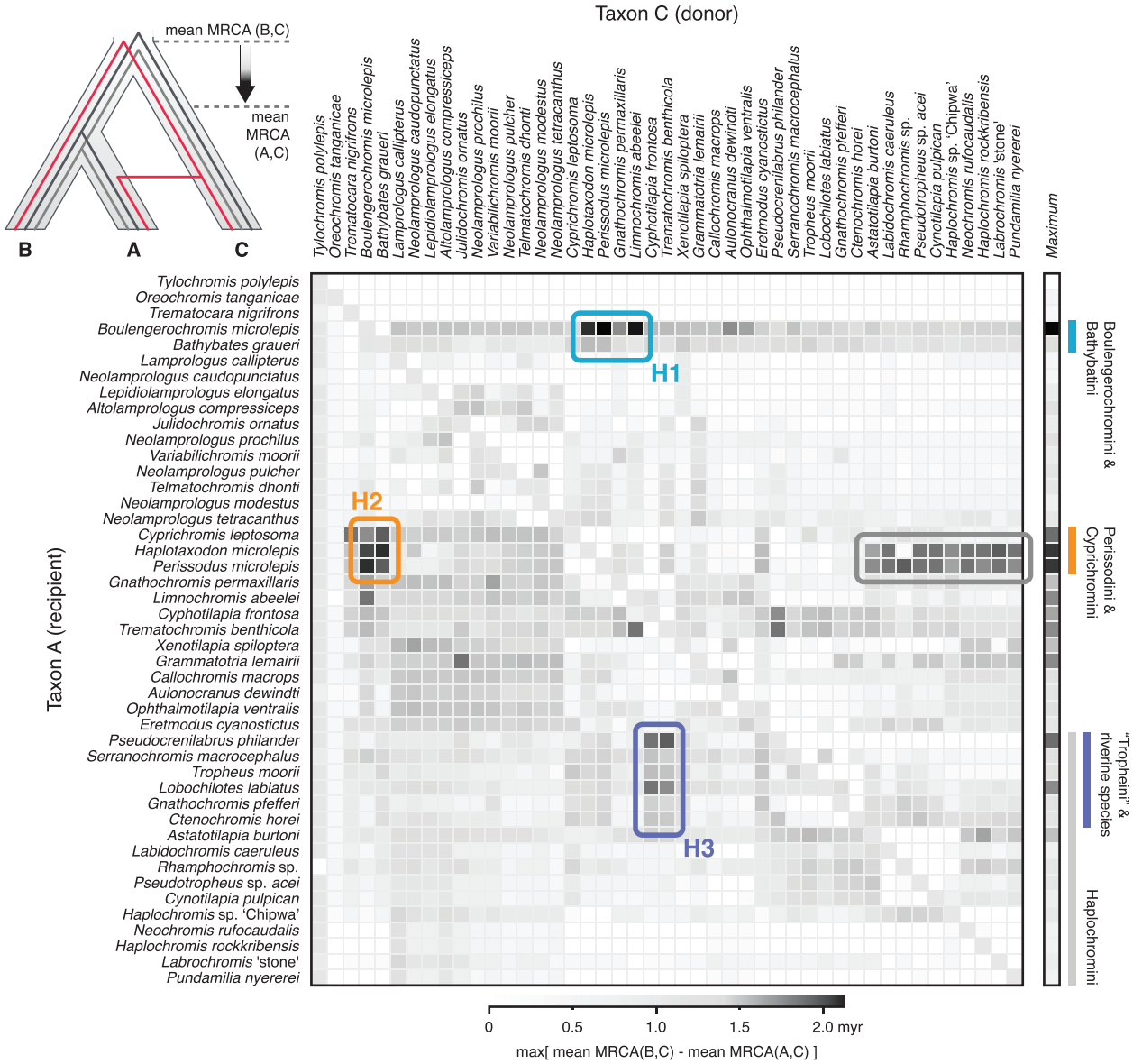

FIGURE 1. Differences in mean MRCA age estimates in three-taxon comparisons. The inset at the top left indicates how introgression from taxon C to taxon A can decrease the mean age estimate for the MRCA of A and C, compared to the mean MRCA age estimate between taxa B and C. In this inset, the species tree is indicated as the light gray shape, with gene trees embedded inside of it. Heat-map cells indicate the maximum age difference between these mean MRCA age estimates, with taxa A and C of the three-taxon comparison according to cell row and column, and any of the remaining species as taxon B. The darkest heat-map cells indicate differences in mean MRCA age estimates of up to 2.13 myr and serve as the basis for three hypotheses of introgression, H1–H3, marked by frames (see text). The frame in gray indicates a fourth possible introgression event between Haplochromini and Perissodini. For each species, the maximum age difference in mean MRCA age estimates, in all possible three-taxon comparisons that include this species as taxon A, is shown in the column to the right of the heat map.

investigate introgression between the Haplochromini and the Perissodini, and instead focussed on testing the first three hypotheses H1–H3.

f-statistics.—The f_4-statistic, introduced by Reich et al. (2009), is a powerful measure to distinguish introgression from incomplete lineage sorting, based on allele frequencies of four populations. With populations A, B, C, and D, and the assumed unrooted population topology (A,B),(C,D), the f_4-statistic is calculated as the product of the difference of allele frequencies between A and B, and between C and D. While this statistic was initially applied to populations of a single species or sister species (Reich et al. 2009), it has been shown to be effective also to detect introgression between species that have diverged several millions of years ago (Martin et al. 2015). We thus calculated the f_4-statistic for sets of four taxa according to our hypotheses of introgression outlined above. To test for introgression between the H-lineage and the Boulengerochromini or Bathybatini (hypothesis H1), we assumed an unrooted

topology (A,B),(C,D), where A was one of the two taxa *Oreochromis tanganicae* and *Trematocara nigrifrons*, B was *Boulengerochromis microlepis* or *Bathybates graueri*, C was either *Perissodus microlepis* or *Limnochromis abeelei*, and D was a member of Lamprologini, either *Lepidiolamprologus elongatus* or *Neolamprologus prochilus*. In these comparisons, the two members of Lamprologini were chosen as this tribe, which is not included in the H-lineage, showed no or very weak signals of introgression into the Boulengerochromini and Bathybatini based on MRCA age estimates, in contrast to the Limnochromini and Perissodini. To test for introgression between the Boulengerochromini or Bathybatini and the Perissodini and Cyprichromini (H2), we repeated this test, again with either *O. tanganicae* or *T. nigrifrons* as taxon A, *B. microlepis* or *B. graueri* as taxon B, and now *P. microlepis* or *Cyprichromis leptosoma* as taxon C, and *Grammatotria lemairii* or *Ophthalmotilapia ventralis* as taxon D. Finally, introgression between Cyphotilapiini and a clade combining Lake Tanganyika Haplochromini with riverine lineages (H3) was assessed with one of the two members of Ectodini as species A, a member of Cyphotilapiini as species B, either *Pseudocrenilabrus philander* or *Lobochilotes labiatus* as species C, and *A. burtoni* or *Haplochromis* sp. 'Chipwa' as species D.

In the absence of introgression, the f_4-statistic is expected to be zero, regardless of whether incomplete lineage sorting is present or not. Thus, introgression between one of the two species A and B and one of the species C and D can be inferred if the f_4-statistic is significantly different from zero (Reich et al. 2009). Whether or not this is the case is usually assessed on the basis of standard errors calculated through a block jackknife procedure. The use of jackknife standard errors for confidence interval estimation assumes that the underlying data is normally distributed, however, this may often not be the case for the f_4-statistic, especially with more divergent species-level allele frequency data. This is because with more divergent populations, a larger numbers of single nucleotide polymorphisms (SNPs) will be fixed within populations but different between them. As a result, the f_4-statistic will be exactly zero for a large number of SNPs. If jackknife blocks include a large number of linked sites, the per-block f_4-statistic may then also be close to zero more often than assumed under normality.

To use the f_4-statistic as a test of introgression with our population-level multimarker data set, we therefore applied not only a block jackknife procedure but also developed a new approach to assess significance, which does not assume normality and accounts for linkage of genetic variation within markers. To this end, we conducted simulations to evaluate how often the observed f_4-statistic can be reproduced in the absence of introgression, based on incomplete lineage sorting alone. We used the coalescent software fastsimcoal2 v.2.5.2 (Excoffier et al. 2013) to produce sequence data sets for four populations that are similar to the true data set in terms of size and amount of missing data. Simulations were carried out using a wrapper script for fastsimcoal2, and simulation parameters for effective population sizes and divergence times were optimized during a burn-in phase. The burn-in phase was stopped as soon as parameter combinations were found with which the resulting simulated sequence variation matched the observed sequence variation across all markers in the proportion of SNPs that are variable in more than one species and in the proportion of SNPs that are variable within both pairs of species. Subsequent to burn-in, 1000 sets of coalescent simulations were carried out, where in each of these sets sequence data was simulated separately for all 39–40 markers included in a given four species comparison. For each comparison, we used only sites that were biallelic among the four species included in the comparison. We interpreted the observed f_4-statistic as evidence for introgression if less than 5% of the 1000 data sets simulated without introgression produced f_4 values at least as extreme as the observed. Our simulation-based test procedure was implemented in the new software F4, which is available at https://github.com/mmatschiner/F4.

Species-Tree Inference with Reduced Taxon Sets

As some of the species showed stronger signals of introgression than others, we assumed that excluding the species with the strongest signals would produce a largely introgression-free reduced taxon set, for which the multispecies coalescent model of *BEAST should be appropriate in a joint analysis of all markers. Based on signals of introgression observed with gene tree MRCA age comparisons and the f_4-statistic, we excluded the following taxa from species-tree inference with *BEAST: *B. microlepis*, *B. graueri*, members of Perissodini and Cyprichromini, the riverine lineages *Pseudocrenilabrus* and *Serranochromis*, as well as all Lake Tanganyika haplochromines except *A. burtoni* and *Haplochromis* sp. 'Chipwa' (i.e., only the "Tropheini"; Takahashi 2003). After excluding all individuals of these species, our data set contained population-level data for 34 remaining species. As we did for gene-tree analyses with *BEAST, we again used the Yule process as the species-tree prior and an HKY model of sequence evolution without among-site rate variation and base frequencies as empirically observed. Time calibration was again based on a strict molecular clock model with marker-specific clock rates and on the same two time constraints based on McMahan et al. (2013) for the ages of Tylochromini and Oreochromini. As alternatives to these time constraints, we also used younger constraints for the same divergence events, based on the timeline of Friedman et al. (2013), in a separate set of analyses. According to the estimates of Friedman et al. (2013), we constrained both the divergence between Tylochromini and Austrotilapiini and the divergence between Oreochromini and Austrotilapiini with lognormal prior distributions, parameterized with means (in real space) 24.283 and 13.914 Ma and standard deviations 0.085 and 0.1 Ma, respectively. Per-branch effective population sizes were estimated as part of all

analyses, however, for computational reasons, the data set was reduced to include maximally four randomly selected phased sequences per species for each marker. To assess robustness of results, random sequence selection was performed twice, and the two resulting data sets were used for independent sets of *BEAST analyses, in each case with both sets of time constraints. For each data set, we performed four replicate *BEAST runs with 4 billion MCMC steps per replicate, of which the first 40% were discarded as burn-in. These analyses were conducted on the Abel cluster of the University of Oslo and required about 80 CPU days per run replicate.

To test how the application of the multispecies coalescent model influences the estimated species-tree topology and timeline of cichlid diversification, we also performed BEAST analyses with concatenated marker sets for the 34 putatively introgression-free species. It has been argued that concatenation is a special case of the multispecies coalescent model in which all gene trees are constrained to be identical in topology and branch lengths (Edwards et al. 2016). Thus, by comparing results based on concatenation and the multispecies coalescent model, we effectively assess how the release of the constraint of identical gene trees for all markers affects the species-tree estimate. We performed Bayesian analyses of concatenated marker sets using the same model of sequence evolution, molecular clock, and speciation as for species-tree inference with the multispecies coalescent model of *BEAST and the time constraints of McMahan et al. (2013). For analyses of concatenated marker sets, we removed within-species sequence variation by random selection of a single sequence per marker and species. This step was repeated twice to assess the effect of stochasticity in the random sequence selection. For the two generated data sets, we conducted five replicate BEAST analyses, each using 500 million MCMC generations, of which we discarded the first 20% as burn-in.

Run convergence in all BEAST analyses was evaluated by ESS values above 200 for nearly all model parameters (and above 100 for all parameters), and by visual inspection of parameter traces within and between replicates. Posterior tree samples of run replicates were combined to produce maximum clade credibility (MCC) species trees with the software TreeAnnotator v.2.1.3 (Bouckaert et al. 2014), with node heights set to mean age estimates.

Constrained Gene-Tree Inference Based on the Species Tree

We now used the species tree based on the putatively introgression-free reduced taxon set to obtain more reliable estimates of gene trees, which were to be used for likelihood inference of introgression (see below). These new gene-tree analyses were performed separately for each of 35 markers, again excluding five markers for which no sequence data for *T. polylepis* were available. Constrained gene trees were inferred with a new set of *BEAST analyses, with additional constraints according to the previously inferred species tree based on the timeline of McMahan et al. (2013). These constraints were placed on the age and topology of 15 clades that were supported with Bayesian posterior probabilities (BPP) of at least 0.99 in the species tree. To constrain the age of these clades according to the previously inferred species tree, we used lognormal prior distributions fitted to the distribution of posterior age estimates from the previous analysis. However, the topology of these clades could not simply be constrained as monophyletic with respect to all other taxa, as the new set of gene-tree analyses included additional species (see below) that might fall into these clades. To allow inclusion of additional taxa within these clades, we used "CladeConstraint" priors from the Sampled Ancestors (Gavryushkina et al. 2014) package for BEAST. These priors allow specifications of in-group and outgroup taxa for a clade, and all taxa not listed in either of the two categories are allowed to be included in one or the other. Data sets used for these analyses included maximally four sequences per species, for the same set of 34 species that were included in the inference of the putatively introgression-free reduced species tree, and for one of the three species *B. microlepis*, *P. microlepis*, and *P. philander*. Three sets of analyses were conducted so that each included only one of these three species, and within-species sequence data was again selected at random, which was repeated twice to produce two equivalent data sets for separate analyses. Thus, for each of the 35 markers, we used six data sets that differed slightly in species sets and selected sequences per species. Per marker, three replicate *BEAST analyses were carried out for each of the six data sets, with the same settings as for the previous species-tree analyses, except that only 500 million MCMC generations were used for each analysis replicate. After run convergence was assessed, we merged the posterior tree distributions for each set of three run replicates. For each marker, and for each of the data sets used for this marker, we produced summary trees with only one tip per species by randomly removing all but one individual per species from each tree of the posterior distribution before generating MCC gene trees. Since visual inspection showed that for each marker and species set, the two MCC gene trees resulting from repeated random selection of within-species sequence data were highly congruent, we generated joint MCC gene trees from the combined posterior tree distributions of the two analyses.

Likelihood-Based Tests for Introgression

We used the inferred sets of constrained gene trees to test the three hypotheses of introgression, H1–H3, in a likelihood framework. For the three sets of taxa that each included one of the three species with strong signals of introgression (*B. microlepis*, *P. microlepis*, and *P. philander*), the inferred MCC gene trees were used jointly to assess the phylogenetic position and the most probable source of introgression into this species. We used function "InferNetwork_ML" (Yu et al. 2014) of the program PhyloNet v.3.5.6 (Than

et al. 2008) to compare the likelihoods of the set of gene trees under incomplete lineage sorting alone, and after adding a single introgression edge to the species tree. For computational reasons, and because we expected introgression into *B. microlepis*, *P. microlepis*, and *P. philander* to originate from basal branches or stem lineages of tribes rather than from their crowns, we reduced all gene trees to single representatives of tribes for these analyses. As representatives of tribes, of which our taxon set included multiple species, we chose the lamprologine *N. prochilus*, the ectodine *O. ventralis*, the limnochromine *G, permaxillaris*, the cyphotilapiine *Trematochromis benthicola*, and the haplochromine *Pundamilia nyererei*, due to low amounts of missing data for these taxa and their nested positions in the respective tribes. After further removing 2 of the 35 gene trees due to missing data for Trematocarini, all 33 remaining gene trees included a tip for each of 10 tribes. To disregard topologically uncertain nodes in gene tree's, we applied a gene tree's support threshold of BPP $\geq$ 0.9 in all PhyloNet analyses (option "-b"). Ten replicate analyses were performed for all sets of gene trees (option "-x") to assess convergence, and branch lengths and inheritance probabilities of introgression edges were optimized after the topology search (option "-po").

Results

Detection of Putatively Introgressed Lineages

Comparison of mean MRCA age estimates.—Our comparison of MRCA mean age estimates in 35 gene trees showed that divergences in three-taxon comparisons were the least tree-like when members of three groups of Lake Tanganyika cichlids were included. These groups comprised (i) the Boulengerochromini and Bathybatini, (ii) the Perissodini and Cyprichromini, and (iii) Lake Tanganyika Haplochromini and the riverine lineages *Pseudocrenilabrus* and *Serranochromis* (Fig. 1). When *B. microlepis* was included in the comparison as taxon A (see scheme in Fig. 1), the MRCA age estimates between taxa A and C were up to 2.13 myr younger than those between taxa A and B, which was the case in a comparison with *T. nigrifrons* as taxon B and *P. microlepis* as taxon C (Supplementary Table S3 available on Dryad). For the three species involved in this comparison, *B. microlepis* was closest to *T. nigrifrons* in the MCC gene trees of 15 markers, closest to *P. microlepis* in the MCC gene trees of 12 markers, and ancestral to the other two taxa in the MCC gene trees of 5 markers (the trees of the remaining markers did not include one of these species) (Supplementary File S2 available on Dryad). Comparable differences in MRCA age estimates were observed when other members of the Perissodini or Limnochromini were included in the comparison instead of *P. microlepis* (Fig. 1), or when *B. graueri* was used instead of *B. microlepis*.

When members of Cyprichromini and Perissodini were used as taxon A in three-taxon comparisons, the greatest difference between the mean age estimate for the MRCA of taxa B and C and the mean age estimate for the MRCA of taxa A and C was found in the species trio *P. microlepis* (as taxon A), *G. pfefferi* (as taxon B), and *B. microlepis* (as taxon C). The mean MRCA age estimate of *P. microlepis* and each of the other two species was about 2 myr younger than that between them (Supplementary Table S3 available on Dryad). *Perissodus microlepis* was found closer to *G. pfefferi* in the MCC gene trees of 18 markers, and closer to *B. microlepis* in the MCC gene trees of 15 markers, and it was the outgroup to the other two taxa in the remaining 5 MCC gene trees (Supplementary File S2 available on Dryad). Very similar patterns were found when other members of Perissodini and Cyprichromini were used instead of *P. microlepis*, or when *B. microlepis* was replaced by *B. graueri* (Fig. 1).

We further observed large differences between the mean age estimate for the MRCA of taxa B and C and the mean age estimate for the MRCA of taxa A and C in comparisons involving Lake Tanganyika haplochromines, *Pseudocrenilabrus*, or *Serranochromis* as taxon A, together with a representative of the Cyphotilapiini as taxon C (Fig. 1). These differences were most pronounced in a comparison including *P. philander* as taxon A, *Neochromis rufocaudalis* as taxon B, and *T. benthicola* as taxon C, where the mean MRCA age estimate for *T. benthicola*, and *N. rufocaudalis* was about 1.8 myr older than both mean age estimates for MRCA involving *P. philander* (Supplementary Table S3 available on Dryad). Among these three species, *P. philander* was sister to *N. rufocaudalis* in the MCC gene trees of 15 markers and to *T. benthicola* in the MCC gene trees of 12 markers, and *N. rufocaudalis* and *T. benthicola* were most closely related in the remaining 7 MCC gene trees (Supplementary File S2 available on Dryad).

Taken together, these results are consistent with at least three hypothesized introgression events during the early diversification of Lake Tanganyika cichlids: MRCA age differences and gene topology frequencies in three-taxon comparisons involving *B. microlepis* might be explained by introgression from the H-lineage to the Boulengerochromini or to a common ancestor of Boulengerochromini and Bathybatini (hypothesis H1), introgression may further have occurred between a member of the Boulengerochromini and Bathybatini, and an ancestor of Perissodini and Cyprichromini (hypothesis H2), and *Pseudocrenilabrus* or an early representative of Lake Tanganyika haplochromines may have experienced introgression from the Cyphotilapiini (hypothesis H3).

f-statistics.—The f_4-statistic was calculated for a total of 46 four-taxon comparisons, with between 390 and 745 biallelic SNPs that could be extracted from our multimarker data set (Supplementary Table S4 available on Dryad). Between 2.0 and 5.4% of these SNPs were variable within both pairs of sister taxa, indicating gene flow between lineages that are separated by millions of years. Most of the calculated f_4 values (39 out of

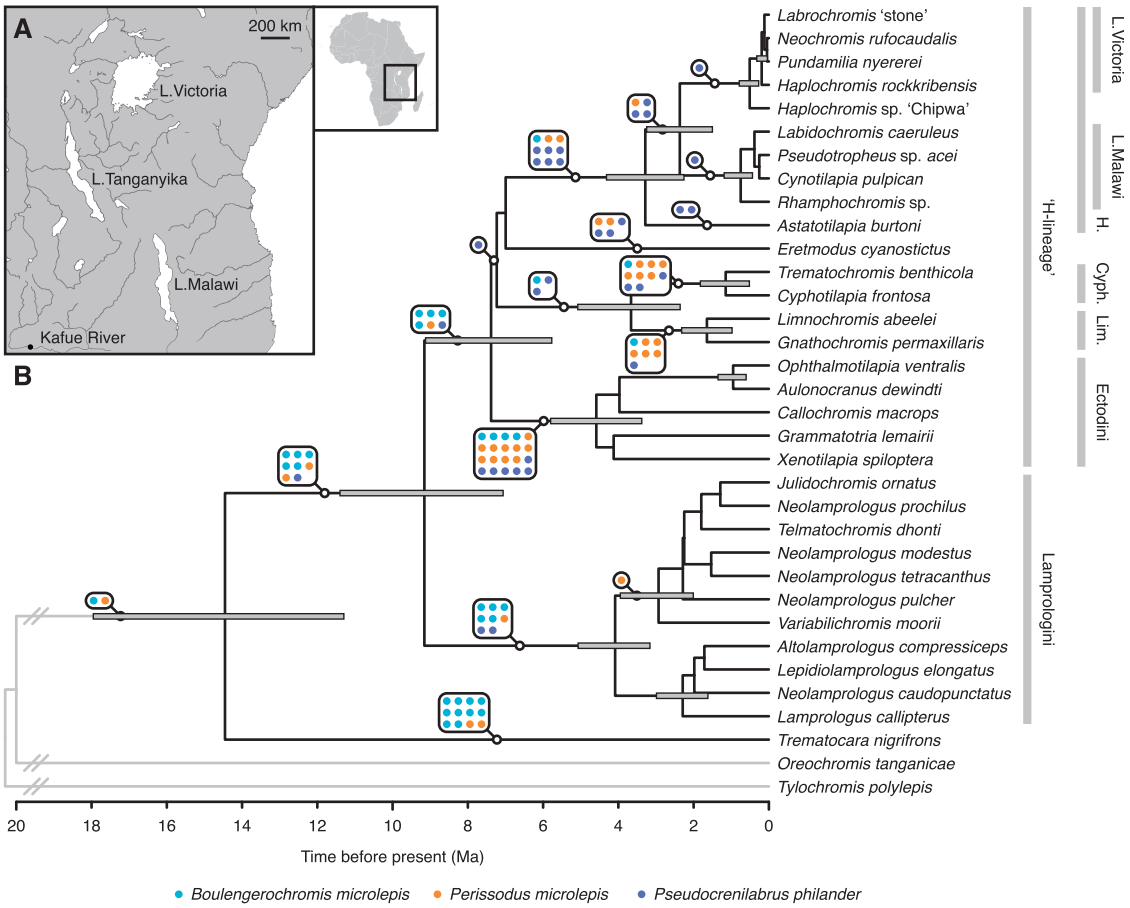

FIGURE 2. Species tree for East African cichlid lineages. A) Map of East Africa showing the three great lakes Tanganyika, Malawi, and Victoria. B) Time-calibrated species tree for a reduced taxon set of 34 East African cichlid species with no or weak signals of introgression, based on 35 markers and inferred using the multispecies coalescent model. Node bars indicate 95% HPD intervals for clades supported with a BPP of 0.99 or higher. Colored dots on branches indicate how often *B. microlepis*, *P. microlepis*, and *P. philander*, respectively, were found to attach to these branches (or to less strongly supported descendent branches) in MCC gene trees resulting from the constrained gene-tree inference. Outgroup branches are not drawn to scale. H., Haplochromini; Cyph., Cyphotilapiini; Lim., Limnochromini.

46) were negative, suggesting gene flow according to the three tested hypotheses of introgression, H1–H3. Based on a block jackknife procedure with a block size of 20 SNPs, f_4 values were significantly different from zero in six comparisons; however, the assumption of normality of jackknife block f_4 values was violated in all of them (Shapiro–Wilks test for normality, $P \leq 0.02$). This indicates that a block jackknife procedure may not be suitable to assess significance of the f_4-statistic. By simulating data sets without introgression that were comparable to the observed data set in terms of SNP number and variation, we found that per four-taxon comparisons, between 11% and 49% of these data sets produced f_4 values as extreme or more extreme than the empirically observed f_4 value (see Supplementary Table S4 available on Dryad).

Species-Tree Inference with Reduced Taxon Sets

Recall, that to test for robustness of phylogenetic inference, we performed all species-tree analyses with two different data sets resulting from random selection of sequences as representatives of within-species variation. For each set of analyses, these two data sets always produced two near-identical phylogenetic estimates of the interrelationships between Lake Tanganyika cichlid tribes. Rather than discussing these two phylogenies separately, we thus combined the two posterior tree samples for each of these analyses to produce joint MCC phylogenies as our best species-tree estimates. We also focus our discussion of results on species trees inferred with time constraints according to McMahan et al. (2013; Fig. 2), but present results were based on analyses with time constraints according to Friedman et al. (2013) in Figure 3, and Supplementary Figure S3, and Supplementary File S2 (available on Dryad). Note that the choice of time constraints did not affect the MCC topology of the species tree except in one weakly supported node (BPP = 0.50) within Ectodini (Fig. 2; Supplementary Fig. S3 available on Dryad), and that branch lengths were between 18.9% and 53.8% (mean = 38.8%) shorter in the analyses with time constraints according to Friedman et al. (2013).

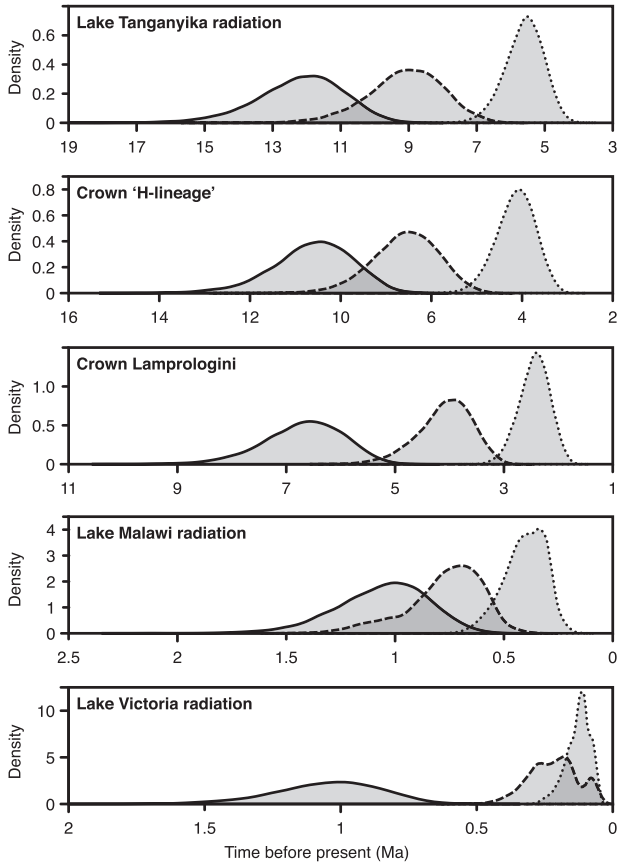

FIGURE 3. Age estimates for five selected clades of East African cichlids. Posterior age distributions for five selected clades, resulting from Bayesian species-tree analyses based on concatenation (solid black line) or the multispecies coalescent model (with time constraints according to McMahan et al. 2013: dashed black line; with time constraints according to Friedman et al. 2013: dotted black line).

Species-tree inference for the 34 taxa with no or only weak signs of introgression produced a strongly supported topology (mean BPP = 0.83) (Fig. 2; Supplementary File S2 available on Dryad). In this species tree, the Trematocarini are recovered as the sister group to all other Tanganyikan cichlid tribes (BPP = 1.0), with the exception of the Tylochromini and Oreochromini that invaded the lake secondarily (Salzburger et al. 2002a). We found strong support (BPP = 1.0) for reciprocal monophyly of the Lamprologini and a clade combining all members of the H-lineage. The monophyly of each tribe is equally well supported, as is the sister group relationship between Limnochromini and Cyphotilapiini, and the monophyly of both the Lake Victoria and Lake Malawi cichlids (BPP = 1.0). Overall, the tree topology corroborates the relationships recently identified with a concatenated analysis of a similar marker set (Meyer et al. 2015), but is unable to resolve the sequence of divergences between the Ectodini, the combined Cyphotilapiini and Limnochromini, the Eretmodini and the Haplochromini. Bayesian time calibration indicates that these four groups have diverged near-instantaneously around 7 Ma (95% highest posterior density intervals [HPD] 9.1–5.3 Ma; Fig. 2).

The use of concatenated marker sets instead of the multispecies coalescent model led to a number of differences in both the topology and the estimated divergence dates (Supplementary Fig. S4; Supplementary File S2 available on Dryad). Most notably, *Eretmodus cyanostictus* appeared as the sister to a clade combining the Haplochromini, Ectodini, Limnochromini, and Cyphotilapiini in the species tree based on concatenation (BPP = 1.0), but nested within these tribes in analyses using the multispecies coalescent model. With the exception of outgroup splits, divergences between Lake Tanganyika tribes were always older when using concatenation (Fig. 3).

The sequence of divergences between the Haplochromini, Eretmodini, Ectodini, and the combined Limnochromini and Cyphotilapiini appeared far less rapid based on concatenation, with mean age estimates between 10.5 (95% HPD 12.7–9.3 Ma) and 8.7 Ma (95% HPD 10.7–6.9 Ma), as compared to 7.4 (95% HPD 9.1–5.8 Ma) to 7.0 Ma (95% HPD 8.7–5.4 Ma) in the species tree based on the multispecies coalescent model. Similarly, the divergence of Lamprologini and the H-lineage was estimated at 12.2 Ma (95% HPD 14.7–9.7 Ma) based on concatenation, but at 9.2 Ma (95% HPD 11.4–7.1 Ma) when using the multispecies coalescent model. Relative differences between age estimates were greatest for the two young radiations of Lake Malawi and Lake Victoria. The MRCA of the four Lake Malawi species included in our taxon set was estimated at 1.1 Ma (95% HPD 1.5–0.7 Ma) based on concatenation, but around 0.7 Ma (95% HPD 1.2–0.4 Ma) in analyses using the multispecies coalescent model. As our taxon set includes cichlids of both the 'mbuna' group and of the genus *Rhamphochromis*, the divergence of these species is likely equivalent to the onset of the Lake Malawi radiation (Joyce et al. 2011; Genner et al. 2015; McGee et al. 2016). The first divergence among the four Lake Victoria species of our taxon set was estimated at 0.9 Ma (95% HPD 1.3–0.6 Ma) in concatenation-based analyses, but only around 191 thousand years ago (ka) (95% HPD 322–40 ka) in the species tree based on the multispecies coalescent model (Fig. 3). With genera *Neochromis*, *Haplochromis*, *Labrochromis*, and *Pundamilia* included in our taxon set, this divergence event is also likely to be synonymous with the origin of the superflock endemic to Lake Victoria (Verheyen et al. 2003), excluding older Lake Kivu haplochromines.

Effective population sizes (N_e) estimated with the multispecies coalescent model ranged between 3.6×10^4 and 8.1×10^5, assuming a mean generation times of 3 years for cichlid fishes (Malinsky et al. 2015). Notably, two out of the three overall lowest effective population sizes of internal branches were found in the two lineages leading to the Lake Malawi ($N_e = 7.5 \times 10^4$; 95% HPD 1.4×10^4–1.4×10^5) and Lake Victoria ($N_e = 1.4 \times 10^5$; 95% HPD 2.2×10^4–2.7×10^5) radiations.

In contrast, high effective population sizes were inferred for the basal lineages of the Lake Tanganyika radiation, with estimates of 6.1×10^5 (95% HPD 2.6–9.8×10^5) for the branch leading to the H-lineage, and 3.8×10^5 (95% HPD 2.4–5.3×10^5) in the stem lineage of Lamprologini, and 5.1×10^5 (95% HPD 2.9–7.3×10^5) in the common ancestor of these two groups.

Constrained Gene-Tree Inference based on the Species Tree

Adding species-tree constraints to the inference of individual gene trees greatly improved the species-level node support of these gene trees (mean BPP = 0.64, compared to mean BPP = 0.43 in unconstrained gene-tree analyses). The addition of species putatively involved in introgression to the taxon set resulted in very different attachment points for *B. microlepis*, *P. microlepis*, and *P. philander* in individual constrained gene trees (Fig. 2; Supplementary File S2 available on Dryad). *Boulengerochromis microlepis* was found in a sister-group position to *T. nigrifrons* in ten gene trees, but also appeared more closely related to the Ectodini, the H-lineage, the Lamprologini, or a clade combining the H-lineage and Lamprologini in a total of 18 constrained gene trees. In contrast, *P. microlepis* was most often recovered as the sister of the Ectodini (in 10 constrained gene trees), but appeared closely related to the Cyphotilapiini (6 constrained gene trees) and Limnochromini (5 constrained gene trees), and also diverged from basal branches in several of the constrained gene trees. Attachment points for *P. philander* were mostly found within the Haplochromini (3 constrained gene trees), but the species also appeared as the sister of the Ectodini (6 constrained gene trees), the Eretmodini (3 constrained gene trees), or the Cyphotilapiini (three constrained gene trees). All attachment points of these three species are shown in Figure 2.

Likelihood-Based Tests for Introgression

Maximum likelihood network topologies were recovered consistently in at least 9 out of 10 analysis replicates for each set of analyses with the software PhyloNet, indicating run convergence. All resulting maximum likelihood network topologies were completely congruent with the previously inferred species tree (Fig. 2), except that *E. cyanostictus* appeared closer to the Cyphotilapiini and Limnochromini than to the Haplochromini when *P. philander* was included in the taxon set. The addition of single introgression edges to the tree improved the likelihood by 28.5, 14.2, and 2.9 log units when *B. microlepis*, *P. microlepis*, and *P. philander*, respectively, were included in the taxon set. *B. microlepis* was recovered as the sister lineage of the clade combining the H-lineage and Lamprologini, and appears to have received introgression from the common ancestor of Cyphotilapiini and Limnochromini, with an estimated inheritance probability of 24% (Fig. 4a).

Perissodus microlepis, on the other hand was found as the sister of Cyphotilapiini and Limnochromini, and a part of its genome appears to have introgressed from the common ancestor of H-lineage and Lamprologini, with a slightly lower inferred inheritance probability of 21% (Fig. 4b). Thus, the two introgression edges inferred for *B. microlepis* and *P. microlepis* are opposing each other, which could be taken as indication that gene flow between the common ancestor of H-lineage and Lamprologini and the common ancestor of Cyphotilapiini and Limnochromini was bidirectional and occurred through hybridization between a member of Boulengerochromini (or a shared ancestor with Bathybatini) and a member of Perissodini (or a shared ancestor with Cyprichromini). Finally, *P. philander* was recovered as the sister lineage of other Haplochromini, and appears to have received introgression from Cyphotilapiini, with an estimated inheritance probability of 9% (Fig. 4c). Thus, we find strong support ($\geq$ 14.2 log-likelihood units) for hypotheses H1 and H2 of introgression between the H-lineage and the Boulengerochromini, and moderate support (2.9 log-likelihood units) for hypothesis H3 of introgression from the Cyphotilapiini to Lake Tanganyika Haplochromini ("Tropheini").

Discussion

A Refined Species-Tree Topology for Lake Tanganyika Cichlids

With hundreds of species of fish, mollusks, and crustaceans found nowhere else on earth, Lake Tanganyika is a textbook example of the evolution of a complex ecosystem in isolation (Salzburger et al. 2014). In particular, the endemic fauna of cichlid fishes has received considerable scientific attention including phylogenetic estimates using molecular sequence data. Earlier mtDNA-based phylogenetic hypotheses for Lake Tanganyika cichlids—sometimes in combination with a small number of nuclear markers—were characterized by a partial lack of resolution (especially with respect to relationships between tribes) and low support values for many nodes, as well as inconsistencies between studies (Kocher et al. 1995; Salzburger et al. 2002b; Clabaut et al. 2005; Day et al. 2008; Koblmüller et al. 2008; Muschick et al. 2012). The rapid development of sequencing techniques leading to the availability of multiple independent molecular markers, in association with novel algorithms, promises to improve species-tree estimates (Delsuc et al. 2005; Lemmon and Lemmon 2013).

In a recent study (Meyer et al. 2015), we presented a well-supported phylogenetic hypothesis of the relationships between Lake Tanganyika cichlid tribes based on 42 nuclear loci using a concatenation approach. The strategy of concatenating multilocus data has some limitations, though. For example, in situations of prevalent hybridization or incomplete lineage sorting, a concatenation approach might not be able to redraw

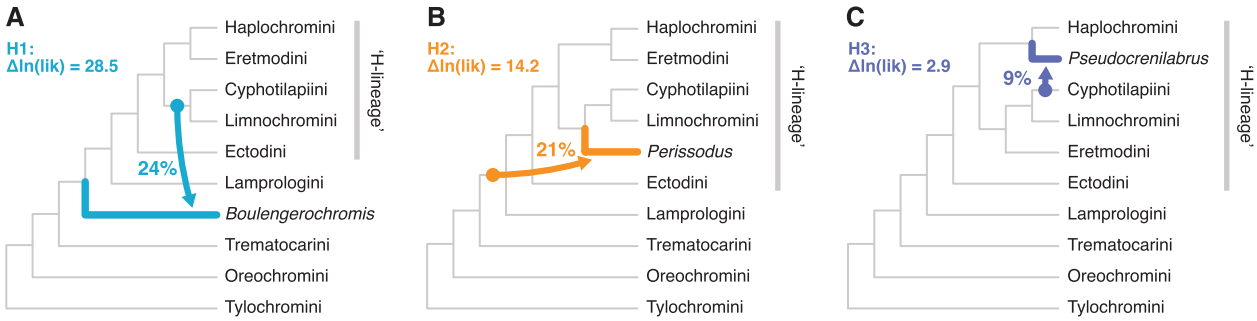

FIGURE 4. The three inferred introgression events between East cichlid tribes. Introgression events were inferred with the likelihood framework implemented in PhyloNet, according to hypotheses H1 (a), H2 (b), and H3 (c). Arrows indicate the direction of introgression events, and percentages given represent inheritance probabilities estimated for introgression edges. The improvement in likelihood gained by addition of this introgression edge is indicated. Note that the donor branch of inferred introgression events can be older than the recipient branch if introgression occurred indirectly through species not included in the phylogeny.

the correct species history. Incomplete lineage sorting is particularly common in rapidly diverging lineages, such as in adaptive radiations, due to the relatively small number of generations between speciation events (Suh et al. 2015). The degree of incomplete lineage sorting also increases with effective population sizes, which are likely to be high for cichlids in Lake Tanganyika (Koblmüller et al. 2006), and could thus lead to deep coalescence events and gene trees that are incongruent to the actual species tree.

Gene-tree discordance originating from stochastic processes such as incomplete lineage sorting can be accounted for with coalescent-based approaches, which has been demonstrated on theoretical grounds (Degnan and Salter 2005; Heled and Drummond 2010), and based on simulated (Leaché and Rannala 2011) and empirical data (Linkem et al. 2016). However, simulation-based studies have also suggested that even coalescent-based species-tree estimates can be misleading when gene flow is present between lineages (Leaché et al. 2014). To overcome this problem, species included in the analysis should be carefully selected to exclude taxa that are affected by introgression. Species-tree estimates can be further improved when multiple individuals per lineage are sampled, as this increases the number of coalescent events used to calculate population sizes (Heled and Drummond 2010; Knowles and Kubatko 2010).

In the present study, we thus increased allele sampling for the taxon set used in Meyer et al. (2015) (we here use up to 16 specimens per species), and applied the multispecies coalescent model to obtain a refined estimation of the species tree of Lake Tanganyika cichlids. To meet the requirements of the multispecies coalescent model, we first determined the set of species with the strongest signals of putative introgression, and excluded these from species-tree analyses. For comparative reasons, we also reconstructed the tree from the same data set but using the concatenation approach.

In the resulting "introgression-free" species tree based on the multispecies coalescent model (Fig. 2), the only representative of the Trematocarini (*T. nigrifrons*) was placed as the sister lineage to a clade formed by the Lamprologini and all H-lineage tribes. Within the H-lineage, Ectodini was recovered as the sister group to a clade formed by Limnochromini and Cyphotilapiini, together with Eretmodini and Haplochromini. The positioning of Eretmodini within the H-lineage—as sister lineage to Haplochromini—is in agreement with previous phylogenetic hypotheses based on allozyme- (Nishida 1991), AFLP (Amplified fragment length polymorphism)- (Weiss et al. 2015), and multilocus data (Friedman et al. 2013; Meyer et al. 2015; McGee et al. 2016), but stands in contrast to inferences based on mtDNA, which placed Eretmodini outside the H-lineage (Kocher et al. 1995; Clabaut et al. 2005; Day et al. 2008; Muschick et al. 2012). In our concatenated analysis, Eretmodini appeared at yet another position, at the base of the H-lineage (Supplementary Fig. S4 available on Dryad; also see Fig. 2b of Meyer et al. 2015). Based on the observed cytonuclear discordance and the variable placement of Eretmodini, it has previously been suggested that this tribe originated from an ancient hybridization event involving a Lamprologini-like ancestor and a member of the H-lineage (Clabaut et al. 2005; Meyer et al. 2015; Weiss et al. 2015). Our approach of identifying introgression based on MRCA ages in three-taxon comparisons (Fig. 1), f-statistics, and likelihood analyses did not retrieve a hybrid signature for *E. cyanostictus*. At the same time, the placement of Eretmodini as sister lineage to Haplochromini is only relatively weakly supported (Fig. 2), calling for additional analyses using genome-wide data and denser taxon sampling to firmly place this tribe in the phylogeny of East African cichlids.

Overall, however, the application of the multispecies coalescent model, subsequent to the exclusion of the species showing the strongest signals of introgression, resulted in a refined species-tree estimate for Lake Tanganyika cichlids compared to our previous study (Meyer et al. 2015). For example, we recovered strong support for the monophyly of Limnochromini and Cyphotilapiini (see also Friedman et al. 2013; McGee et al. 2016), which both occur in the deep-water habitat of Lake Tanganyika (Coulter 1991). Two other branches remain weakly supported—the one placing Ectodini at the base of the H-lineage and the above-mentioned branch placing Eretmodini as sister lineage

to Haplochromini. The apparent lack of phylogenetic resolution among these branches in our analyses with the multispecies coalescent model (Fig. 2) is likely to reflect the rapidity of lineage formation at the base of the H-lineage (see also Salzburger et al. 2002a; Koblmüller et al. 2004; Duftner et al. 2005; Koblmüller et al. 2008). While the basal branches within the H-lineage receive higher support values based on the concatenation model (Supplementary Fig. S4 available on Dryad), these may be misleading, as concatenation has been shown to overestimate branch support and potentially produce incorrect species-tree estimates (Kubatko and Degnan 2007; Ogilvie et al. 2016; Linkem et al. 2016). In contrast, using the multispecies coalescent model with only tens of loci may produce more accurate species trees than concatenation with far larger data sets (Ogilvie et al. 2016).

The drawback of our strategy to produce an "introgression-free" species-tree estimate is that it comes on the expense of the exclusion of certain taxa (the putative recipients of introgression). The resulting species tree is therefore robust, yet incomplete. Nevertheless, there are ways to interpret the placement of the taxa that were excluded from species-tree analyses due to signals of introgression. For example, under the assumption that the true phylogenetic signal is stronger than the signal resulting from introgression, the frequencies of the attachment of a lineage to a given branch in the individual gene trees can be considered as democratic vote for its phylogenetic position (indicated by dots on branches in Fig. 2). We exemplify this for the three taxa with the strongest signals of introgression. Based upon the inspection of individual gene trees, *B. microlepis* is suggested to be a sister lineage of *T. nigrifrons* (as e.g., Salzburger et al. 2002a; Clabaut et al. 2005; Meyer et al. 2015), or, alternatively, more closely related to the common ancestor the Lamprologini and the H-lineage (as in Weiss et al. 2015). *Perissodus microlepis* shows affinities to the ancestor of Ectodini (see, e.g., Kocher et al. 1995; Clabaut et al. 2005) or is placed in a clade together with Limnochromini and Cyphotilapiini (as observed in Salzburger et al. 2002a; Friedman et al. 2013). Finally, the riverine haplochromine *P. philander* is most often placed near the base of the Haplochromini clade (see, e.g., Salzburger et al. 2002a; Clabaut et al. 2005; Santos et al. 2014; Meyer et al. 2015), but also appears frequently as sister lineage to the Ectodini, which has not been observed previously.

Our strategy to separate phylogenetic signals of species divergences from those of incomplete lineage sorting and introgression assumes that the latter two processes are the only sources of gene-tree discordance. However, it has been shown that with increasing size of the data set used for phylogenetic inference, the degree of systematic errors due to violations of the model of sequence evolution, can also increase and produce misleading phylogenetic signal (Rodríguez-Ezpeleta et al. 2007). These model violations may arise, for example, from changes in site-specific substitution rates over time (heterotachy; Pagel and Meade 2008), from among-species heterogeneity in sequence composition (Rodríguez-Ezpeleta et al. 2007), or from GC-content variation among taxa (Jarvis et al. 2014). If and to what degree these model violations affect our data set, with possible consequences for phylogenetic inference, has not been tested explicitly. However, it is likely that misleading phylogenetic signals due to these model violations are stronger on longer timescales (as in Jarvis et al. 2014), suggesting that systematic errors may have little effect on the inference of the relatively recent diversification of Lake Tanganyika cichlids.

A number of alternative approaches, which have not been explored here, are available for species-tree inference while taking hybridization into account. A phylogenetic model similar to that used in the "InferNetwork_ML" method of PhyloNet (Yu et al. 2014) is also implemented in the software PhyloNetworks (Solís-Lemus and Ané 2016), as a method named "SNaQ" (Species Networks applying Quartets). In contrast to "InferNetwork_ML", however, "SNaQ" does not calculate the full likelihood of this model, but a pseudolikelihood after decomposition of the phylogeny into four-taxon subsets, which allows faster calculations and therefore the application to larger data sets, at the cost of comparability of different networks (Solís-Lemus and Ané 2016). The option to infer networks based on their pseudolikelihood (calculated from three-taxon subsets) instead of their full likelihood has also been added to the software PhyloNet, as a method named "InferNetwork_MPL" (Yu and Nakhleh 2015). A model for the joint inference of incomplete lineage sorting and hybridization is also available in the software STEM-hy (Kubatko 2009). However, in this comparatively early implementation, the species topology is assumed to be known and the hybridization event is considered to lead to hybrid speciation (Mallet 2007), which apparently is far less common than introgressive hybridization, in which genetic material is exchanged between lineages without the production of a hybrid species (Schumer et al. 2014).

Application of the Multispecies Coalescent Model Improves Age Estimates for East African Radiations

The timeline of diversification estimated with the multispecies coalescent model was consistently younger than that obtained with concatenated sequences for all markers. This observation is in agreement with the differences between the models employed by the two approaches. While the concatenation model implicitly assumes that all gene trees are identical to each other as well as to the species tree (Edwards et al. 2016), the multispecies coalescent model accounts for genetic divergence that predates speciation. As a result, concatenation is likely to result in overestimated speciation times, and the degree of this overestimation depends on the ancestral effective population sizes (McCormack et al. 2011). With large population sizes, as are commonly found in adaptive radiations of fishes

(Won et al. 2005), the difference in age estimates obtained with the two approaches can be on the order of millions of years (Colombo et al. 2015; Ogilvie et al. 2016). As all previous fossil-based phylogenetic time calibrations of the East African cichlid diversification were based on the concatenation model (e.g., Genner et al. 2007; Schwarzer et al. 2009; Friedman et al. 2013; McMahan et al. 2013), our timeline estimated with the multispecies coalescent model is likely to provide more realistic divergence dates for cichlid radiations in lakes Tanganyika, Malawi, and Victoria.

Lake Tanganyika radiation.—Even though seismic reflection data in combination with extrapolations of Pleistocene sedimentation rates suggest that Lake Tanganyika originated around 12–9 Ma (Cohen et al. 1993, 1997), previous concatenation-based age estimates for tribes endemic to this lake are far older. Regardless of whether an assumed Gondwanan vicariance or the cichlid fossil record were used for divergence-time analyses, estimates of Genner et al. (2007) imply that at least 10 cichlid lineages colonized Lake Tanganyika independently but left no traces of their existence outside of the lake. Using calibrations based on the phylogeny of Azuma et al. (2008) and the fossil *Oreochromis* †*lorenzoi* (Carnevale et al. 2003), Schwarzer et al. (2009) estimated the MRCA of Lamprologini and the H-lineage between 20.4 and 10.6 Ma, with a mean age estimate of 15.4 Ma. This estimate is comparable to our concatenation-based age estimate for the same node (12.2 Ma), but is likely to predate the origin of Lake Tanganyika. In contrast, our analyses with the multispecies coalescent model, time calibrated according to McMahan et al. (2013), suggest that Lamprologini and the H-lineage diverged around 9.2 Ma, which could thus represent the first divergence event within Lake Tanganyika (Fig. 3). This timeline is, therefore, consistent with a colonization history of Lake Tanganyika that requires two to four colonization events from lineages that subsequently went extinct outside the lake (Trematocarini, the common ancestor of Lamprologini and the H-lineage, and possibly the ancestors of Boulengerochromini and Bathybatini) (Salzburger et al. 2002a). According to estimates based on the multispecies coalescent model in combination with time constraints derived from Friedman et al. (2013), even the divergence of Trematocarini (and thus possibly also those of Boulengerochromini and Bathybatini) could have occurred within the lake (around 8.7 Ma; Supplementary Fig. S3 available on Dryad), which would allow the possibility of a single austrotilapiine colonization of Lake Tanganyika.

Within the H-lineage, we identified a sequence of rapid splitting events with the multispecies coalescent model that began with the separation of Ectodini around 7.4 Ma (based on time constraints derived from McMahan et al. 2013) and ended with the divergence of Eretmodini and Haplochromini less than 400,000 years later. In contrast, estimates for these divergence events are stretched out over almost 2 myr when inferred with the concatenation model, which is indicative of a high degree of incomplete lineage sorting and thus discordance of individual gene trees. Our coalescent-based age estimates for these lineages coincide with the topographic depression of the northern Lake Tanganyika basin (8–7 Ma; Cohen et al. 1993; Lezzar et al. 1996), suggesting that the main diversification of Lake Tanganyika tribes could have been triggered by ecological opportunity (see, e.g., Wagner et al. 2012) in this newly formed lake basin.

Lake Malawi radiation.—Lake Malawi has existed perennially for 4.5–4.0 myr (Ring and Betzler 1995), which is consistent with most phylogenetic age estimates for its largely endemic cichlid radiation (Genner et al. 2007; Schwarzer et al. 2009; Friedman et al. 2013; Loh et al. 2013; McMahan et al. 2013). However, throughout its history, Lake Malawi has experienced severe lake-level fluctuations with lowstands that had considerable impact on its cichlid fauna due to habitat reduction and eutrophication (Salzburger et al. 2014; Lyons et al. 2015). Our age estimate based on the multispecies coalescent model and time constraints derived from McMahan et al. (2013) (749 ka) nearly coincides with a shift in hydroclimate regimes during the Mid-Pleistocene Transition (~800 ka) that led to a wetter climate and less frequent lake-level fluctuations (Lyons et al. 2015). These changes could have promoted diversification of Lake Malawi cichlid fishes due to increased ecological opportunity and habitat stability.

Lake Victoria radiation.—While the Lake Victoria basin originated around 400 ka (Johnson et al. 1996; Salzburger et al. 2014), the radiation of the endemic Lake Victoria superflock (*sensu* Verheyen et al. 2003, excluding Lake Kivu species) has long been thought to be younger than 200 ka, based on mitochondrial mutation rate estimates and palaeogeographic considerations (Meyer et al. 1990; Johnson et al. 1996; Verheyen et al. 2003). However, phylogenetic time calibrations have so far failed to reproduce these young age estimates (Elmer et al. 2009; Wagner et al. 2012; Rabosky et al. 2013). Though our timeline based on concatenation suggests an early origin of the Lake Victoria superflock around 913 ka, mean age estimates based on the multispecies coalescent model support a much younger origin (191–160 ka; Fig. 3; Supplementary Fig. 3 available on Dryad) in agreement with previous studies (Meyer et al. 1990; Johnson et al. 1996; Verheyen et al. 2003). In line with this observation, MCC gene trees of Lake Victoria cichlids coalesce much earlier (6.1–0.7 Ma) than their MCC species tree in analyses with the multispecies coalescent model, which indicates that most of the genetic variation present in cichlids of the Lake Victoria superflock pre-existed in the common ancestor. Our estimate of the effective population size of this ancestor ($N_e = 143,000$) is small compared to those of other internal branches but nevertheless suggests a rather large founding population of the Lake Victoria superflock.

This could indicate that Lake Victoria was colonized gradually through previously existing river connections (see Seehausen et al. 2003; Salzburger et al. 2005) rather than by singular dispersal events. It should be noted, however, that effective population size estimation with the multispecies coalescent model assumes an absence of population structure and introgression, and that a violation of these assumptions may result in overestimated population sizes.

Our phylogenetic time calibration was based on secondary constraints taken from the timelines of global cichlid diversification in McMahan et al. (2013) and Friedman et al. (2013), and therefore relies on the accuracy of these timelines. According to age estimates of McMahan et al. (2013), Tylochromini and Oreochromini originated around 42 and 23 Ma, respectively, which is in agreement with the earliest fossil records of both tribes, cf. *Tylochromis* from the Egyptian Jebel Qatrani Formation (35.1–33.8 Ma; Murray 2002, 2004), and *Oreochromis* †*martyini* from the Ngorora Formation of Kenya (12.0–9.3 Ma; VanCouvering 1982; Murray and Stewart 1999). On the other hand, the younger age estimates of Friedman et al. (2013) partially disagree with the fossil record of cichlids, as several clades (e.g., Tylochromini and Hemichromini) appear younger than their first fossil occurrences (Chen et al. 2014; Musilová et al. 2015). In another recent study on the divergence times of cichlid fishes, Matschiner et al. (2016) used a new model for Bayesian clade age estimation in combination with a large number of teleost fossil constraints and obtained relatively older estimates for the ages of Tylochromini (77.0–60.5 Ma) and Oreochromini (63.6–47.0 Ma). However, for reasons discussed in Matschiner et al. (2016), these ages could have been overestimated and should be considered as upper boundaries rather than an accurate timeline. Therefore, we suggest that age estimates based on the intermediate timeline of McMahan et al. (2013) may provide the most realistic picture of East African cichlid diversification to date.

Introgression in Lake Tanganyika Cichlids

The exchange of genetic material between species is a common phenomenon in adaptive radiations (see, e.g., Berner and Salzburger 2015), ranging from ongoing gene flow between recently diverged species to more ancient introgression events (The Heliconius Genome Consortium 2012; Jones et al. 2012; Lamichhaney et al. 2015). Likewise, recent and more ancient hybridization events are well documented for East African cichlid fishes (Lake Tanganyika: Koblmüller et al. 2007, 2010; Sturmbauer et al. 2010; Weiss et al. 2015; Lake Malawi: Joyce et al. 2011; Genner and Turner 2012; Lake Victoria: Keller et al. 2013; admixture across watersheds: Schwarzer et al. 2012; Loh et al. 2013). Moreover, it has been suggested that hybridization might promote adaptive radiation in the first place (Grant and Grant 1992; Seehausen 2004) via the creation of novel combinations of parental alleles leading to increased genetic diversity and novel phenotypes in the offspring, upon which natural and sexual selection can act (Hedrick 2013; Arnold 2015). In some situations, for example, after the colonization of novel habitats, hybrids might be better adapted for the exploitation of new ecological niches (see, e.g., Rieseberg et al. 2003; Seehausen 2004; Stelkens and Seehausen 2009; Abbott et al. 2013; Keller et al. 2013; Seehausen et al. 2014).

Here, we applied a novel approach based on MRCA ages in three-taxon comparisons, f-statistics, and likelihood analyses to detect introgression between members of different cichlid tribes in Lake Tanganyika. Our analyses uncovered three past introgression events that left behind particularly strong signatures in the genomes of the descendant lineages (Figs. 1 and 4). The most strongly supported case of past introgression involves the ancestor of *B. microlepis* (or its common ancestor with Bathybatini) as recipient, and an early member of Perissodini or Limnochromini as donor. In a second event, *B. microlepis* (or an earlier member of Boulengerochromini or Bathybatini) served as donor, whereas the common ancestor of Perissodini and Cyprichromini was identified as the most likely recipient. Since both instances involve—in inverted positions—a similar set of recipient and donor lineages, it is possible that they reflect a single introgression event responsible for the reciprocal exchange of substantial portions of genetic material between a more ancient lineage of cichlids (the ancestor of Boulengerochromini or of Boulengerochromini and Bathybatini) and a member of the more derived H-lineage (see Supplementary Fig. S5 available on Dryad, for an illustration of this scenario). Third, there is evidence for introgression from an early member of Cyphotilapiini into an ancestor of the Lake Tanganyika haplochromines and the closely related riverine genera *Pseudocrenilabrus* and *Serranochromis* (see also Weiss et al. 2015). Within the clade combining Haplochromini of Lake Tanganyika with the two riverine lineages, the strongest signals of introgression were observed in *P. philander*, closely followed by the Lake Tanganyika endemic *L. labiatus*. While we here used the riverine *P. philander* in all analyses of introgression as the representative of this clade (as we consistently chose the taxon with the strongest signals in each case), it seems likely that the introgression event affecting this clade took place within Lake Tanganyika and thus subsequent to its secondary colonization by haplochromines. The presence of introgression signals not only in Lake Tanganyika haplochromines but also in riverine lineages could then be explained by their divergence subsequent to introgression, or by further genetic exchange between these closely related species (Loh et al. 2013).

The introgression events inferred in our study are largely congruent with those observed by Weiss et al. (2015) based on a large AFLP data set. In line with our findings, Weiss et al. (2015) detected signals of introgression between ancient Tanganyika tribes (i.e., Boulengerochromini, Bathybatini, Hemibatini, and Trematocarini) and the H-lineage. Furthermore, Weiss et al. (2015) found that support for monophyly of a clade

combining all H-lineage tribes except Haplochromini (the "ancient mouthbrooders," a clade not recovered in our analysis) increased after excluding members of Cyphotilapini. This was taken as evidence for genetic exchange between Cyphotilapini and other tribes outside of the "ancient mouthbrooders," in agreement with our detection of introgression between Cyphotilapini and Haplochromini or related riverine lineages. Thus, despite disagreement in the inferred relationships between tribes and the use of very different data sets, the study of Weiss et al. (2015) and the present study potentially identified traces of the same introgression events among Lake Tanganyika tribes. Taken together, it thus appears that the adaptive radiation of cichlid fishes in Lake Tanganyika is characterized by several introgression events at different times rather than a major event producing a hybrid swarm at the onset of the radiation, as has been suggested for other adaptive radiations (Seehausen 2004; Hudson et al. 2010), including the cichlids from lakes Victoria (Seehausen et al. 2003) and Malawi (Joyce et al. 2011).

What could have triggered introgressive hybridization between distantly related cichlid lineages (i.e., at the level of different tribes) in the confines of Lake Tanganyika, and under which circumstances and environmental conditions such rare hybridization events might have taken place, remains a matter of speculation. There are, however, arguments that render past gene exchange between *B. microlepis* and a member of the H-lineage, and between Cyphotilapiini and Haplochromini, not entirely improbable. *Boulengerochromis microlepis*, the only member of Boulengerochromini, is unusual in various aspects. Originally assigned to Tilapiini (Poll 1986), *B. microlepis* was put into its own tribe because of its distinctiveness with respect to any other cichlid species known (Takahashi 2003). *Boulengerochromis microlepis* is the largest cichlid in the world, reaching an adult size of ~80 cm, and it forms breeding pairs for unusually long periods of time, until the offspring has almost reached adult size (Konings 2015). Our results suggest that *B. microlepis* is characterized by a combination of genetic material from a pre-Tanganyikan cichlid lineage and a member of the endemic H-lineage. Thus, gene exchange with lake-adapted lineages might have contributed to the survival of the ancestor of *B. microlepis* after its colonization of this novel environment.

The Tanganyikan representatives of the Haplochromini, the "Tropheini" (Takahashi 2003), are also somewhat unusual compared to other members of that tribe. For example, most "Tropheini" species are much less sexually dimorphic than for example, the haplochromines of the adaptive radiations in lakes Victoria and Malawi (Fryer and Iles 1972; Salzburger et al. 2006), and some of the Tanganyikan haplochromines become unusually large (up to 40 cm in length and 3 kg in size the case of *Petrochromis* sp. 'giant'). Interestingly, there are morphological similarities between the Tanganyikan haplochromines and the Cyphotilapiini. In fact, both genera currently included in Cyphotilapiini, *Cyphotilapia* and *Trematochromis*, have initially been grouped into the Haplochromini based on morphological grounds (Poll 1986), and were later suggested to form their own, respective monotypic tribes because of their phenotypic distinctiveness (Takahashi 2003). The only representative of the latter genus (*T. benthicola*) was, originally, even placed into the same genus as the Tanganyikan haplochromine *Ctenochromis horei* and has only recently been established as a member of the Cyphotilapiini based on molecular data (Muschick et al. 2012).

In the light of our new results suggesting ancient gene flow from Cyphotilapiini into Haplochromini, the morphological resemblance between the members of these tribes can be reinterpreted as being due to past introgression. Our data place the introgression event from Cyphotilapiini to Haplochromini near the common ancestor of the "Tropheini" and riverine lineages (*Pseudocrenilabrus* and *Serranochromis*), while other, more derived, Haplochromini of Lake Malawi and Lake Victoria do not seem to be affected. This suggests that the radiation of the haplochromines in Lake Tanganyika started out from a riverine ancestor that invaded the lake secondarily and hybridized with a lake-adapted species. This scenario is in line with the idea that hybridization can trigger adaptive radiation (Seehausen 2004), for example, through transgressive segregation (Stelkens and Seehausen 2009). Unlike in the haplochromine adaptive radiations in lakes Victoria and Malawi, for which precursory hybridization between exclusively riverine lineages has been proposed (Seehausen et al. 2003; Joyce et al. 2011), the Tanganyikan haplochromines appear to be the product of a riverine and a lake-adapted form. Recent studies have emphasized the contribution of riverine species in the formation of the lacustrine East African cichlid radiations, both by introducing new lineages and through introgressive hybridization (Joyce et al. 2011; Loh et al. 2013). As the present study includes only few riverine taxa, the influence of these species could not be explored in detail. To extend the picture of introgression mediated by riverine taxa, it will thus be important to also include representatives of other closely related riverine groups (e.g., *Orthochromis*, *Serranochromis*, or *Astatoreochromis*) in future analyses.

The here reported cases of past introgression between different cichlid tribes in Lake Tanganyika have in common an asymmetric fate with respect to taxonomic diversity of the two lineages involved in the exchange of genetic material. Each event brought forth an ancestor of a lineage that subsequently diversified into an array of species (Perissodini and Cyprichromini: 16 species; Haplochromini: ~30 species in Lake Tanganyika), while leaving behind another lineage with low species diversity (Cyphotilapiini: 3 species; Boulengerochromini: 1 species). What might have caused this asymmetry in speciation rates after gene exchange, and how introgression is connected to diversification overall in Lake Tanganyika cichlids, should be in the focus of future studies. In this context, it is interesting to note that we did not

find evidence for intertribal introgression events that involve a member of the most species-rich cichlid tribe in Lake Tanganyika, Lamprologini, although—within this tribe—introgression and hybridization appear to be common (Salzburger et al. 2002b; Sturmbauer et al. 2010). A possible explanation for the "genomic demarcation" of lamprologines from other cichlid lineages in Lake Tanganyika is their distinctive breeding mode: Lamprologines are substrate spawners with only very subtle, if any, phenotypic differences between males and females, whereas all but one (*B. microlepis*) of the remaining Tanganyikan cichlid species are mouth-brooders featuring, in many cases, a pronounced sexual (color) dimorphism (Konings 2015).

Taken together, we uncovered several instances of past intertribal introgression in Lake Tanganyika. These involve ancestors of tribes that, until today, could not have been placed firmly into the phylogeny of East African cichlids, suggesting that their chimeric nature is responsible for the difficulties in resolving their respective placements in previous analyses based on both molecular and morphological data. Our identification of multiple introgression events in Lake Tanganyika cichlids is in line with expectations regarding the occurrence of hybridization in adaptive radiation and corroborates the importance of between-species genetic exchange for the evolution of biodiversity.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.62qn5.

Funding

This study was supported by grants from the Swiss National Science Foundation (SNF) and the European Research Council (ERC; StG "Intergenadapt" and CoG "CICHLID~X").

Acknowledgments

We would like to thank Marco Colombo, Bernd Egger, Adrian Indermaur, Moritz Muschick, Frauke Münzel, Fabrizia Ronco, and Anya Theis for assistance during fieldwork, and Ole Seehausen for providing tissue samples; Brigitte Aeschbach and Nicolas Boileau for help with lab work; Christof Wunderlin and Georges Wigger from Microsynth for 454 sequencing; and Alexei Drummond, Thomas Marcussen, Caleb McMahan, and Huw Ogilvie for discussion. We further thank the editor Thomas Near, the Associate Editor Laura Kubatko, Lukas Rüber, and one anonymous reviewer for helpful comments on the manuscript. Calculations were performed at the sciCORE (http://scicore.unibas.ch/) scientific computing core facility at University of Basel, and on the Abel Cluster (www.hpc.uio.no/), owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR), and operated by the Department for Research Computing at USIT, the University of Oslo IT-department.

References

Abbott R., Albach D., Ansell S., Arntzen J.W., Baird S.J.E., Bierne N., Boughman J., Brelsford A., Buerkle C.A., Buggs R., Butlin R.K., Dieckmann U., Eroukhmanoff F., Grill A., Cahan S.H., Hermansen J.S., Hewitt G., Hudson A.G., Jiggins C.D., Jones J., Keller B., Marczewski T., Mallet J., Martinez-Rodriguez P., Möst M., Mullen S., Nichols R., Nolte A.W., Parisod C., Pfennig K., Rice A.M., Ritchie M.G., Seifert B., Smadja C.M., Stelkens R., Szymura J.M., Väinölä R., Wolf J.B.W., Zinner, D. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Arnold M.L. 2015. Divergence with genetic exchange. Oxford: Oxford University Press.

Azuma Y., Kumazawa Y., Miya M., Mabuchi K., Nishida M. 2008. Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. BMC Evol. Biol. 8:215.

Bell M.A., Foster, S.A. 1994. The evolutionary biology of the Threespine Stickleback. Oxford, Oxford University Press.

Berner D., Salzburger, W. 2015. The genomics of organismal diversification illuminated by adaptive radiations. Trends Genet. 31:491–499.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLOS Comput. Biol. 10:e1003537.

Brawand D., Wagner C.E., Li Y.I., Malinsky M., Keller I., Fan S., Simakov O., Ng A.Y., Lim Z.W., Bezault E., Turner-Maier J., Johnson J., Alcazar R., Noh H.J., Russell P., Aken B., Alföldi J., Amemiya C., Azzouzi N., Baroiller J.-F., Barloy-Hubler F., Berlin A., Bloomquist R., Carleton K.L., Conte M.A., D'Cotta H., Eshel O., Gaffney L., Galibert F., Gante H.F., Gnerre S., Greuter L., Guyon R., Haddad N.S., Haerty W., Harris R.M., Hofmann H.A., Hourlier T., Hulata G., Jaffe D.B., Lara M., Lee A.P., MacCallum I., Mwaiko S., Nikaido M., Nishihara H., Ozouf-Costaz C., Penman D.J., Przybylski D., Rakotomanga M., Renn S. C.P., Ribeiro F.J., Ron M., Salzburger W., Sanchez-Pulido L., Santos M.E., Searle S., Sharpe T., Swofford R., Tan F.J., Williams L., Young S., Yin S., Okada N., Kocher T.D., Miska E.A., Lander E.S., Venkatesh B., Fernald R.D., Meyer A., Ponting C.P., Streelman J.T., Lindblad-T. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381.

Carnevale G., Sorbini, C., Landini W. 2003. *Oreochromis lorenzoi*, a new species of tilapiine cichlid from the Late Miocene of central Italy. J. Vert. Paleontol. 23:508–516.

Chen W.-J., Santini F., Carnevale G., Chen J.-N., Liu S.-H., Lavoué S. Mayden R.L. 2014. New insights on early evolution of spiny-rayed fishes (Teleostei: Acanthomorpha). Front. Mar. Sci. 1:53.

Clabaut C., Salzburger W., Meyer A. 2005. Comparative phylogenetic analyses of the adaptive radiation of Lake Tanganyika cichlid fish: nuclear sequences are less homoplasious but also less informative than mitochondrial DNA. J. Mol. Evol. 61:666–681.

Cohen A.S., Soreghan M.J., Scholz, C.A. 1993. Estimating the age of formation of lakes: an example from Lake Tanganyika, East African Rift system. Geology 21:511–514.

Cohen A.S., Lezzar K.E., Tiercelin J.J., Soreghan M. 1997. New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. Basin Research 9:107–132.

Colombo M., Damerau M., Hanel R., Salzburger W., Matschiner M. 2015. Diversity and disparity through time in the adaptive radiation of Antarctic notothenioid fishes. J. Evol. Biol. 28:376–394.

Coulter G.W. 1991. The benthic fish community. In: Coulter, G.W.; editor, Lake Tanganyika and its Life, London: Oxford University Press, p. 151–199.

Coyne J.A., Orr H.A. 2004. Speciation. Sunderland (MA): Sinauer Associates, Inc.

Day J.J., Cotton J.A., Barraclough T.G. 2008. Tempo and mode of diversification of Lake Tanganyika cichlid fishes. PLOS One 3:e1730.

Degnan J.H., Salter L.A. 2005. Gene tree distributions under the coalescent process. Evolution 59:24–37.

Delsuc F., Brinkmann, H., Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6:361–375.

Duftner N., Koblmüller S., Sturmbauer C. 2005. Evolutionary relationships of the Limnochromini, a tribe of benthic deepwater cichlid fish endemic to Lake Tanganyika, East Africa. J. Mol. Evol. 60:277–289.

Dunz A.R., Schliewen U.K. 2013. Molecular phylogeny and revised classification of the haplotilapiine cichlid fishes formerly referred to as "*Tilapia*". Mol. Phylogenet. Evol. 68:64–80.

Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leaché A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.

Elmer K.R., Reggio C., Wirth T., Verheyen E., Salzburger W., Meyer A. 2009. Pleistocene desiccation in East Africa bottlenecked but did not extirpate the adaptive radiation of Lake Victoria haplochromine cichlid fishes. Proc. Natl. Acad. Sci. USA 106:13404–13409.

Excoffier L., Dupanloup I., Huerta-Sachez E., Sousa V.C., Foll M. 2013. Robust demographic inference from genomic and SNP data. PLOS Genet. 9:e1003905.

Friedman M., Keck B.P., Dornburg A., Eytan R.I., Martin C.H., Hulsey C.D., Wainwright P.C., Near T.J. 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. Proc. R. Soc. B 280:20131733.

Fryer G., Iles T.D. 1972. The Cichlid Fishes of the Great Lakes of Africa: their biology and evolution. Scotland: Oliver and Boyd Press.

Gavryushkina A., Welch D., Stadler T., Drummond A.J. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. PLOS Comput. Biol. 10:e1003919.

Genner M.J., Ngatunga B.P., Mzighani S., Smith A., Turner G.F. 2015. Geographical ancestry of Lake Malawi's cichlid fish diversity. Biol. Lett. 11:20150232.

Genner M.J., Seehausen O., Lunt D.H., Joyce D.A., Shaw P.W., Carvalho G.R., Turner G.F. 2007. Age of cichlids: new dates for ancient lake fish radiations. Mol. Biol. Evol. 24:1269–1282.

Genner M.J., Turner G.F. 2012. Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. Mol. Biol. Evol. 29:195–206.

Glor R.E. 2010. Phylogenetic insights on adaptive radiation. Annu. Rev. Ecol. Evol. Syst. 41:251–270.

Grant P.R., Grant B.R. 1992. Hybridization of bird species. Science 256:193–197.

Grant P., Grant B. 2008. How and why species multiply: the radiation of Darwin's finches. Princeton (NJ): Princeton University Press.

Hasegawa M., Kishino H., Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Hedrick P.W. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol. 22:4606–4618.

Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27:570–580.

Holder M.T., Anderson J.A., Holloway A.K. 2001. Difficulties in detecting hybridization. Syst. Biol. 50:978–982.

Holland B.R., Benthin S., Lockhart P.J., Moulton V., Huber K.T. 2008. Using supernetworks to distinguish hybridization from lineage-sorting. BMC Evol. Biol. 8:202.

Hudson A.G., Vonlanthen P., Seehausen O. 2010. Rapid parallel adaptive radiations from a single hybridogenic ancestral population. Proc. R. Soc. B 278:58–66.

Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S., Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., daFonseca R.R., Li J., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldón T., Capella-Gutiérrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C.V., Lovell P.V., Wirthlin M., Schneider M., P.C., Prosdocimi F., Samaniego J.A., VargasVelazquez A.M., Alfaro-Núez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu B., Wu K., Xiao J., Yinqi X., Zheng, Q., Zha. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320–1331.

Johnson T., Scholz C., Talbot M., Kelts K., Ricketts R., Ngobi G., Beuning K., Ssemmanda I., McGill J. 1996. Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. Science 273:1091–1093.

Joly S., McLenachan P.A., Lockhart P.J. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am. Nat. 174:E54–E70.

Jones F.C., Grabherr M.G., Chan Y.F., Russell P., Mauceli E., Johnson J., Swofford R., Pirun M., Zody M.C., White S., Birney E., Searle S., Schmutz J., Grimwood J., Dickson M.C., Myers R.M., Miller C.T., Summers B.R., Knecht A.K., Brady S.D., Zhang H., Pollen A.A., Howes T., Amemiya C., Baldwin J., Bloom T., Jaffe D.B., Nicol R., Wilkinson J., Lander E.S., DiPalma F., Lindblad-Toh K., Kingsley D.M. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Joyce D.A., Lunt D.H., Genner M.J., Turner G.F., Bills R., Seehausen O. 2011. Repeated colonization and hybridization in Lake Malawi cichlids. Curr. Biol. 21:R108–R109.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Keller I., Wagner C.E., Greuter L., Mwaiko S., Selz O.M., Sivasundar A., Wittwer S., Seehausen O. 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Mol. Ecol. 22:2848–2863.

Knowles L.L., Kubatko L.S., eds. 2010. Estimating species trees: practical and theoretical aspects. Hoboken (NJ): Wiley-Blackwell.

Koblmüller S., Duftner N., Sefc K.M., Aibara M., Stipacek M., Blanc M., Egger B., Sturmbauer C. 2007. Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika – the result of repeated introgressive hybridization. BMC Evol. Biol. 7:7.

Koblmüller S., Egger B., Sturmbauer C., Sefc K.M. 2010. Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. Mol. Phylogenet. Evol. 55:318–334.

Koblmüller S., Salzburger W., Sturmbauer C. 2004. Evolutionary relationships in the sand-dwelling cichlid lineage of Lake Tanganyika suggest multiple colonization of rocky habitats and convergent origin of biparental mouthbrooding. J. Mol. Evol. 58:79–96.

Koblmüller S., Schliewen U.K., Duftner N., Sefc K.M., Katongo C., and Sturmbauer, C. 2008. Age and spread of the haplochromine cichlid fishes in Africa. Mol. Phylogenet. Evol. 49:153–169.

Koblmüller S., Sefc K.M., Duftner N., Warum M., Sturmbauer C. 2006. Genetic population structure as indirect measure of dispersal ability in a Lake Tanganyika cichlid. Genetica 130:121–131.

Kocher T.D., Conroy J.A., McKaye K.R., Stauffer J.R., Lockwood S.F. 1995. Evolution of NADH dehydrogenase subunit 2 in East African cichlid fish. Mol. Phylogenet. Evol. 4:420–432.

Kocher T.D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. Nat. Rev. Genet. 5:288–298.

Konings A. 2015. Tanganyika cichlids in their natural habitat. 3rd ed. (TX): Cichlid Press, El Paso.

Kubatko L.S. 2009. Identifying hybridization events in the presence of coalescence via model selection. Syst. Biol. 58:478–488.

Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

Lamichhaney S., Berglund J., Almén M.S., Maqbool K., Grabherr M., Martinez-Barrio A., Promerová M., Rubin C.-J., Wang C., Zamani N., Grant B.R., Grant P.R., Webster M.T., Andersson L. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371–375.

Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.

Leaché A.D., Harris R.B., Rannala B., Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63:17–30.

Leaché A.D., Rannala B. 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. 60:126–137.

Lemmon E.M., Lemmon A.R. 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. 44:99–121.

Lezzar K.E., Tiercelin J.J., deBatist M., Cohen A.S., Bandora T., VanRensbergen P., LeTurdu C., Mifundu W., Klerkx J. 1996. New seismic stratigraphy and Late Tertiary history of the North Tanganyika Basin, East African Rift system, deduced from multichannel and high-resolution reflection seismic data and piston core evidence. Basin Res. 8:1–28.

Li H., Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595.

Linkem C.W., Minin V.N., Leaché A.D. 2016. Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae). Syst. Biol. 65:465–477.

Linz S., Radtke A., von Haeseler A. 2007. A likelihood framework to measure horizontal gene transfer. Mol. Biol. Evol. 24:1312–1319.

Loh Y. H.E., Bezault E., Muenzel F.M., Roberts R.B., Swofford R., Barluenga M., Kidd C.E., Howe A.E., DiPalma F., Lindblad-Toh K., Hey J., Seehausen O., Salzburger W., Kocher T.D., Streelman J.T. 2013. Origins of shared genetic variation in African cichlids. Mol. Biol. Evol. 30:906–917.

Losos J. 2009. Lizards in an evolutionary tree. Berkeley (CA): University of California Press.

Lyons R.P., Scholz C.A., Cohen A.S., King J.W., Brown E.T., Ivory S.J., Johnson T.C., Deino A.L., Reinthal P.N., McGlue M.M., Blome M.W. 2015. Continuous 1.3-million-year record of East African hydroclimate, and implications for patterns of evolution and biodiversity. Proc. Natl. Acad. Sci. USA 112:15568–15573.

Maddison W., Knowles L. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.

Malinsky M., Challis R.J., Tyers A.M., Schiffels S., Terai Y., Ngatunga B.P., Miska E.A., Durbin R., Genner M.J., Turner G.F. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science 350:1493–1498.

Mallet J. 2007. Hybrid speciation. Nature 446:279–283.

Marcussen T., Sandve S.R., Heier L., Spannagl M., Pfeifer M., International Wheat Genome Sequencing Consortium, Jakobsen K.S., Wulff B.B.H., Steuernagel B., Mayer K.F.X., Olsen O.-A. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. Science 345:1250092.

Martin C.H., Cutler J.S., Friel J.P., DeningTouokong C., Coop G., Wainwright P.C. 2015. Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. Evolution 69:1406–1422.

Matschiner M., Musilová Z., Barth J.M.I., Starostová Z., Salzburger W., Steel M., Bouckaert R. 2016. Bayesian phylogenetic estimation of clade ages supports trans-Atlantic dispersal of cichlid fishes. Syst. Biol. Advance access, doi:10.1093/sysbio/syw076:1–59.

Mayr E. 2001. What evolution is. New York (NY): Basic Books.

McCormack J.E., Heled J., Delaney K.S., Peterson A.T., Knowles L.L. 2011. Calibrating divergence times on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. Evolution 65:184–202.

McCune A.R., Lovejoy N.R. 1998. The relative rate of sympatric and allopatric speciation in fishes: tests using DNA sequence divergence between sister species and among clades. In: D. Howard and S. Berloccher, editors, Endless forms: species and speciation, p. 172–185. Oxford: Oxford University Press.

McGee M.D., Faircloth B.C., Borstein S.R., Zheng J., DarrinHulsey C., Wainwright P.C., Alfaro M.E. 2016. Replicated divergence in cichlid radiations mirrors a major vertebrate innovation. Proc. R. Soc. B 283:20151413.

McMahan C.D., Chakrabarty P., Sparks J.S., Smith W.L., Davis M.P. 2013. Temporal patterns of diversification across global cichlid biodiversity (Acanthomorpha: Cichlidae). PLOS ONE 8:e71162.

Meyer A., Kocher T.D., Basasibwaki P., Wilson A.C. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature 347:550–553.

Meyer B.S., Matschiner M., Salzburger W. 2015. A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach. Mol. Phylogenet. Evol. 83:56–71.

Meyer B.S., Salzburger W. 2012. A novel primer set for multilocus phylogenetic inference in East African cichlid fishes. Mol. Ecol. Resour. 12:1097–1104.

Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541–i548.

Murray A.M. 2002. Lower pharyngeal jaw of a cichlid fish (Actinopterygii; Labroidei) from an Early Oligocene site in the Fayum, Egypt. J. Vert. Paleontol. 22:453–455.

Murray A.M. 2004. Late Eocene and early Oligocene teleost and associated ichthyofauna of the Jebel Qatrani Formation, Fayum, Egypt. Palaeontology 47:711–724.

Murray A.M., Stewart K.M. 1999. A new species of tilapiine cichlid from the Pliocene, Middle Awash, Ethiopia. J. Vert. Paleontol. 19:293–301.

Muschick M., Indermaur A., Salzburger W. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. Curr. Biol. 22:2362–2368.

Musilová Z., Říčan O., Říčanová Š., Janšta P., Novák O.G.J. 2015. Phylogeny and historical biogeography of trans-Andean cichlid fishes (Teleostei: Cichlidae). Vertebr. Zool. 65:333–350.

Nishida M. 1991. Lake Tanganyika as an evolutionary reservoir of old lineages of East African cichlid fishes: inferences from allozyme data. Experientia 47:974–979.

Nolte A.W., Tautz D. 2010. Understanding the onset of hybrid speciation. Trends Genet. 26:54–58.

Ogilvie H.A., Heled J., Xie D., Drummond A.J. 2016. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. Syst. Biol. Advance Access, doi: 10.1093/sysbio/syv118:1–52.

Pagel M., Meade A. 2008. Modelling heterotachy in phylogenetic inference by reversible-jump Markov chain Monte Carlo. Phil. Trans. R. Soc. B 363:3955–3964.

Poll M. 1986. Classification des Cichlidae du lac Tanganika. Tribus, genres et especes. Académie Royale de Belgique. Classe des Sciences. Mémoires 45:1–163.

Rabosky D.L., Santini F., Eastman J., Smith S.A., Sidlauskas B., Chang J., Alfaro M.E. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat. Commun. 4:1958.

Reich D., Thangaraj K., Patterson N., Price A.L., Singh L. 2009. Reconstructing Indian population history. Nature 461:489–494.

Rieseberg L.H., Raymond O., Rosenthal D.M., Lai Z., Livingstone K., Nakazato T., Durphy J.L., Schwarzbach A.E., Donovan L.A., Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211–1216.

Ring U., Betzler C. 1995. Geology of the Malawi Rift: kinematic and tectonosedimentary background to the Chiwondo Beds, northern Malawi. J. Hum. Evol. 28:7–21.

Roch S., Steel M. 2014. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor. Popul. Biol. 100:56–62.

Rodríguez-Ezpeleta N., Brinkmann H., Roure B., Lartillot N., Lang B.F., Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst. Biol. 56:389–399.

Rokas A., Carroll S.B. 2006. Bushes in the tree of life. PLOS Biol. 4:e352.

Rüber L., Meyer A., Sturmbauer C., Verheyen E. 2001. Population structure in two sympatric species of the Lake Tanganyika cichlid tribe Eretmodini: evidence for introgression. Mol. Ecol. 10:1207–1225.

Salzburger W. 2009. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol. Ecol. 18:169–185.

Salzburger W., Baric S., Sturmbauer C. 2002b. Speciation via introgressive hybridization in East African cichlids? Mol. Ecol. 11:619–625.

Salzburger W., Mack T., Verheyen E., Meyer A. 2005. Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol. Biol. 5:17.

Salzburger W., Meyer A., Baric S., Verheyen E., Sturmbauer C. 2002a. Phylogeny of the Lake Tanganyika cichlid species flock and its

relationship to the central and East African haplochromine cichlid fish faunas. Syst. Biol. 51:113–135.

Salzburger W., Niederstatter H., Brandstatter A., Berger B., Parson W., Snoeks J., Sturmbauer C. 2006. Colour-assortative mating among populations of *Tropheus moorii*, a cichlid fish from Lake Tanganyika, East Africa. Proc. R. Soc. B 273:257–266.

Salzburger W., VanBocxlaer B., Cohen A.S. 2014. Ecology and evolution of the African Great Lakes and their faunas. Annu. Rev. Ecol. Evol. Syst. 45:519–545.

Santos M.E., Braasch I., Boileau N., Meyer B.S., Sauteur L., Böhne A., Belting H.-G., Affolter M., Salzburger W. 2014. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun. 5:1–11.

Schluter D. 2000. The ecology of adaptive radiation. New York (NY): Oxford University Press.

Schmieder R., Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864.

Schumer M., Rosenthal G.G., Andolfatto P. 2014. How common is homoploid hybrid speciation? Evolution 68:1553–1560.

Schwarzer J., Misof B., Tautz D., Schliewen U.K. 2009. The root of the East African cichlid radiations. BMC Evol. Biol. 9:186.

Schwarzer J., Swartz E.R., Vreven E., Snoeks J., Cotterill F.P.D., Misof B.Y., Schliewen U.K. 2012. Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution. Proc. R. Soc. B 279:4389–4398.

Seehausen O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.

Seehausen O. 2006. African cichlid fish: a model system in adaptive radiation research. Proc. R. Soc. B 273:1987–1998.

Seehausen O., Butlin R.K., Keller I., Wagner C.E., Boughman J.W., Hohenlohe P.A., Peichel C.L., Saetre G.-P., Bank C., Brännström Å., Brelsford A., Clarkson C.S., Eroukhmanoff F., Feder J.L., Fischer M.C., Foote A.D., Franchini P., Jiggins C.D., Jones F.C., Lindholm A.K., Lucek K., Maan M.E., Marques D.A., Martin S.H., Matthews B., Meier J.I., Möst M., Nachman M.W., Nonaka E., Rennison D.J., Schwarzer, J., Watson, E.T., Westram, A.M., and Widmer, A. 2014. Genomics and the origin of species. Nat. Rev. Genet. 15:176–192.

Seehausen O., Koetsier E., Schneider M.V., Chapman L.J., Chapman C.A., Knight M.E., Turner G.F., Alphen J.J.M.v., Bills R. 2003. Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. Proc. R. Soc. B 270:129–137.

Simpson G.G. 1953. The major features of evolution. New York: Columbia University Press.

Solís-Lemus C., Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLOS Genet. 12:e1005896.

Stelkens R.B., Seehausen O. 2009. Genetic distance between species predicts novel trait expression in their hybrids. Evolution 63:884–897.

Sturmbauer C., Salzburger W., Duftner N., Schelly R., Koblmüller S. 2010. Evolutionary history of the Lake Tanganyika cichlid tribe Lamprologini (Teleostei: Perciformes) derived from mitochondrial and nuclear DNA data. Mol. Phylogenet. Evol. 57:266–284.

Suh A., Smeds L., Ellegren H. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of Neoavian birds. PLOS Biol. 13:e1002224.

Takahashi T. 2003. Systematics of Tanganyikan cichlid fishes (Teleostei: Perciformes). Ichthyol. Res. 50:367–382.

Takahashi T., Sota T. 2016. A robust phylogeny among major lineages of the East African cichlids. Mol. Phylogenet. Evol. 100:234–242.

Than C., Ruths D., Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinform. 9:322.

The Heliconius Genome Consortium (2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98.

VanCouvering J.A.H. 1982. Fossil cichlid fish of Africa. Special Papers in Paleontology, The Paleontological Association 29:1–103.

Verheyen E., Salzburger W., Snoeks J., Meyer A. 2003. Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. Science 300:325–329.

Wagner C.E., Harmon L.J., Seehausen O. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487:366–369.

Weiss J.D., Cotterill F P.D., Schliewen U.K. 2015. Lake Tanganyika—A 'Melting Pot' of ancient and young cichlid lineages (Teleostei: Cichlidae)? PLOS One 10:e0125043.

Won Y.-J., Sivasundar A., Wang Y., Hey J. 2005. On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence. Proc. Natl. Acad. Sci. USA 102:6581–6586.

Yu Y., Dong J., Liu K.J., Nakhleh L. 2014. Maximum likelihood inference of reticulate evolutionary histories. Proc. Natl. Acad. Sci. USA 111:16448–16453.

Yu Y., Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genom. 16:S10.

Yu Y., Than C., Degnan J.H., Nakhleh L. 2011. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. Syst. Biol. 60:138–149.

Yule G.U. 1925. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. Phil. Trans. R. Soc. B 213:21–87.